Award Number: **W81XWH-11-1-0221**

TITLE: **Biomarkers for pulmonary injury following deployment**

PRINCIPAL INVESTIGATOR: **Dr. Richard Gelinas, PhD**

CONTRACTING ORGANIZATION: **Institute for Systems Biology, 401 Terry Ave. North; Seattle WA 98109-5234**

REPORT DATE: **October 2013**

TYPE OF REPORT: **Final Report**

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2013	Final Report	1 February 2011 – 31 July 2013

4. TITLE AND SUBTITLE

Biomarkers for pulmonary injury following deployment

5a. CONTRACT NUMBER
W81XWH-11-1-0221

5b. GRANT NUMBER
W81XWH-11-1-0221

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Richard Gelinas; Dr. Joseph Brown

Email: Richard.Gelinas@systemsbiology.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Institute for Systems Biology
401 Terry Avenue North
Seattle, WA 98109-5234

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

To understand the risks of lung disease faced by soldiers, USACEHR has studied a rat model of lung disease. After instillation of dust or silica, lung pathology was assessed and proteins and microRNAs that were released into the lung lavage fluid were determined. The released proteins and miRNAs marked the stage-specific pathology. We have since profiled proteins and miRNAs in bronchoalveolar lavage fluid or urine from active duty soldiers with dyspnea who were evaluated in the lung clinic of the San Antonio Military Medical Center as part of the STAMPEDE project under the medical supervision of Dr. Michael J. Morris. We identified groups of markers that are now being correlated with the lung diagnoses made by Dr. Morris. The markers we described may correlate with lung conditions such as asthma and assist in lung disease diagnosis in the future.

15. SUBJECT TERMS
Dust instillation/lung injury/protein biomarker/microRNA biomarker/bronchoalveolar lavage fluid/asthma/urine

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	56	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

Introduction

Over 2.5M military personnel have served in SW Asia since 2002 to the present as part of Operation Iraqi Freedom, Operation Enduring Freedom, or more recently Operation New Dawn. Many of these people were exposed to geologic dust or other airborne particulate matter. After deployment, some military personnel have returned with new symptoms including dyspnea or shortness of breath that required further evaluation and medical attention. The overall goal of our work is to discover molecular signatures or objective biomarkers of lung disease to assist medical authorities in diagnosing these individuals. In collaboration with USACEHR, we studied the patterns of proteins and microRNAs (miRNAs) in bronchoalveolar lavage fluid and serum from a pre-clinical model of lung disease secondary to dust instillation. Our results were reported in the 2012 Annual Report for this contract.

We were able to extend the term of the contract for six months starting in February 2013 so that we could continue to study clinical samples from active duty military personnel with dyspnea that had been collected as part of the 'STudy of Active Duty Military Personnel for Environmental Dust Exposure' (STAMPEDE I) which was created by Dr. Michael J. Morris at the San Antonio Military Medical Center. Here we report the early results from the application of advanced molecular profiling to the clinical samples contributed by the soldiers who enrolled in STAMPEDE I.

By comparing subjects who self-reported with dyspnea to control individuals we established protein profiles of bronchoalveolar lavage fluid (BAL) and urine. We also profiled miRNAs in BAL, serum, and urine. Interestingly, subsets of STAMPEDE subjects were found with groupings of differentially expressed proteins or miRNAs from the lavage data which could be explained if these subjects shared a common diagnosis. Whether they do is currently being evaluated by Dr. Morris and his team.

The molecular profiles we established from the BAL samples from the fifteen control individuals define the state of the normal lung. These independent protein and miRNA biomarkers may be valuable and patentable as a general reference for lung health. For example the healthy lung profiles represent lung 'wellness' which is of interest in the context of personalized medicine but also judging progressive changes in lung health after surgery, recovery from a disease, or treatment with a drug. The ultimate goal will be to describe valid markers for lung disease diagnosis, disease stratification, progression, and response to drug therapy which would provide valuable diagnostic information in the lung clinic.

Body

Methods and Materials

Research participants. Soldiers with post-deployment respiratory symptoms were referred to the STAMPEDE project at the San Antonio Military Medical Center (SAMMC) and attended the pulmonary clinic from 2011 to 2012. Study subjects were evaluated with full pulmonary function studies, radiographic imaging with high resolution chest CT scans, and other testing as clinically appropriate. The standard evaluation included flexible bronchoscopy from which BAL fluid was collected. The BAL fluid was used to study cellularity, flow cytometry and cytokine levels at SAMMC; a portion was sent to the Institute for Systems Biology for miRNA profiling and to Pacific Northwest National Labs for proteomics profiling. Blood serum and urine were also collected from all study subjects and sent to the ISB and PNNL.

Lung fluid and urine sample collection. All patients underwent BAL sampling in the right middle lobe of the lung with 180 cc normal saline instilled (60 cc x 3).

BALF sample preparation for proteome analysis. Samples were thawed and desalted and concentrated with Amicon 3K MWCO spin filters (EMD Millipore, Billerica, MA). First, samples

were concentrated by passing 4 mL of 100 mM NH4HCO3 (buffer) through each filter at 4,000 x g for 40 minutes at 4 C. The volume was adjusted to a total of 4 ml with buffer and centrifuged at 4000 x g, 4 C for 45 minutes. Samples were washed by filling the filter portion with 4 mL of buffer (ensuring resuspension of the sample from the bottom of the filter) and then centrifuged again at speed & temperature as before, except for 1 hour, 15 minutes to ensure the dead volume was reached. Any color changes were noted after concentration. The samples were transferred from the filter portion of each concentrator to a 2.0-mL microcentrifuge tube. The filters were rinsed by adding 100 uL of buffer, vortexing briefly and then using a pipet tip to "wash" the two membranes 3X each. Then the wash sample was combined with the main sample. Next, the volume of each sample was measured and normalized (adjusted) to match the largest volume of the set of samples being processed together. The mass of each sample was calculated by a BCA assay. Urea and DTT were added to final concentrations of 8M and 5 mM, respectively and the samples were reduced and denatured at 60^{o} C for 30 minutes. Iodoacetamide was added to 40 mM and samples were incubated at 37^{o}C for 1 hour to alkylate. The samples were diluted eight-fold with buffer, and $CaCl_2$ was added to 1 mM. Samples were digested with trypsin (in a 1:50 (w:w) ratio of trypsin:protein) at 37^{o} C for 3 hours. Samples were purified on C18 solid phase extraction 'Discovery' columns (Supelco-Sigma-Aldrich, Bellefonte, PA) followed by concentration, assaying protein concentration, and dilution 0.5 ug/uL for MS analysis.

Urine sample preparation for proteome analysis. The urine samples were processed in an automated fashion on an epMotion (Eppendorf, Hauppauge, NY) after loading 500ul of each urine sample into a 1.0 mL 96-well plate and concentrated to dryness. Next, 107 ul of liquid 8M urea was added to the dried urine, vortexed, and then briefly centrifuged. Protein concentration was assayed by the BCA procedure, followed by addition of DTT to 8.3 mM. The plate was vortexed, centrifuged briefly, and incubated for 1 hour with shaking. Iodoacetamide was added

to 36 mM and the plate was incubated at 37° C for 1 hour in the dark with shaking. The samples were then diluted 8-fold with 100 mM NH_4HCO_3 (buffer), $CaCl_2$ was added to 1 mM, and trypsin was added in a 1:50 trypsin:protein (w:w) ratio. The plate was incubated at 37°C for 3 hours with shaking. Samples were purified as above on C18 columns (Agilent, Santa Clara, CA), concentrated, re-assayed for protein concentration and diluted to 0.3 ug/uL for MS analysis.

RPLC separation and MS(/MS) acquisition. The LC system was custom built using two Agilent 1200 nanoflow pumps and one Isco constant pressure capillary pump (Teledyne-Isco, Lincoln, NE), various Valco valves (Valco Instruments Co., Houston, TX), and a PAL autosampler (Leap Technologies, Carrboro, NC). Full automation was made possible by custom software that allows for parallel event coordination and therefore near 100% MS duty cycle through use of two trapping and analytical columns. Reversed-phase columns were prepared in-house by slurry packing 3-µm Jupiter C18 (Phenomenex, Torrence, CA) into 35-cm x 360 µm o.d. x 75 µm i.d fused silica (Polymicro Technologies Inc., Phoenix, AZ) using a 1-cm sol-gel frit for media retention (unpublished PNNL variation of the method of Maiolica et al., 2005). Trapping columns were prepared similarly but using a 4-cm length of 100 µm i.d. fused silica that was fritted on both ends. Mobile phases consisted of 0.1% formic acid in water (A) and 0.1% formic acid acetonitrile (B) operated at 300 nL/min with a gradient profile as follows (min:%B); 0:5, 2:8, 20:12, 70:35, 97:60, 100: 95. Sample injection occurred 40 min prior to beginning the gradient while data acquisition lagged the gradient start and end times by 10 min to account for column dead volume that allowed for the tightest overlap possible in two-column operation. Two-column operation also allowed for columns to be 'washed' (shortened gradients) and re-generated off-line without any cost to duty cycle.

MS analysis was performed using a Velos Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) outfitted with a custom electrospray ionization (ESI) interface. Electrospray emitters were custom made by chemically etching 150 um o.d. x 20 um i.d. fused silica (Kelly et al.,

2006). The heated capillary temperature and spray voltage were 350ºC and 2.2 kV, respectively. Data was acquired for 100 min after a 10 min delay from when the gradient started. Orbitrap spectra (AGC 1x106) were collected from 400-2000 m/z at a resolution of 60k followed by data-dependent HCD MS/MS (collision energy 32%, AGC 5x104) of the ten most abundant ions, excluding single charge states. A dynamic exclusion time of 60 sec was used to discriminate against previously analyzed ions using a -0.55 to 1.55 Da mass window.

Mass spectrometry data analysis. AMT tag results were filtered by for a mass error less than 3 ppm and by STAC for a uniqueness probability score greater than 0.5 and a FDR threshold < 10%. The resulting datasets were $\log_2$ transformed. The optimal normalization algorithm was determined by SPANS (Webb-Robertson et al., 2011) to be a mean center with the rank invariant peptide (RIP) selection having a *p*-value threshold of 0.1 for the BALF and a mean center with the top *L* Order Statistics (LOS) peptide selection having a *p*-value threshold of 0.05 for the urine datasets. The correlation scores were summarized between datasets derived from different individuals as shown in **Figure 1**. A probabilistic principal component analysis, within the pcaMethods package in R, was performed on the datasets containing missing values and the results are presented in **Figure 2**.

Statistical analysis. Hypothesis tests were performed with MSstats (Clough et al., 2009), with missing-action set to remove, to determine statistical differences in protein abundance between control and disease samples. *P*-values were corrected for multiple comparisons using the Benjamini-Hochberg *p*-value adjustment (Benjamini & Hochberg, 1995). For heatmaps, protein abundance vectors were arranged by ascending fold-difference. All statistical tests were performed with R (Team, 2008).

MicroRNA analysis. RNA enriched for miRNA was isolated from 250 microliter aliquots of bronchial alveolar lavage fluid, 150 microliter aliquots of serum, and 250 microliter aliquots of

urine by using the miRNeasy mini kit (Qiagen, cat. 217 004). The concentrations of about 800 human miRNAs were determined by using the NanoString nCounter human miRNA expression assay kit version 2.1 following the manufacturer's instructions (NanoString Technologies, Seattle WA). Data reduction from a first workflow applied to the STAMPEDE patients (n=47) and control individuals (n=15) included the steps of loading raw data files into Excel, combining replicate profiles, and normalizing with the global mean which was calculated separately for the urine, serum, and lavage data sets. Hierarchical cluster analysis was carried out with Multiple Experiment Viewer (Saeed et al., 2003). These results are presented in Figs. 10 through 15 below.

An independent analysis of the data adopted a second workflow leading to the results in Fig. 16. The data was first processed through several steps. First, technical replicates were averaged. Next the data was normalized to minimize lane-by-lane variation by a factor derived from the geometric mean of the positive controls. Next, a background correction was applied, which was based on the mean plus two standard deviations of the negative controls. Finally, the data was again normalized using a factor calculated from the geometric mean of the highly abundant probes. These steps were performed using the R package "NanoStringNorm". Next, the normalized, background corrected data was log2-transformed. Then, using the Bioconductor package 'LIMMA', we identified differentially expressed miRNAs between the STAMPEDE and control groups. P-values from the moderated t-test and fold-changes between groups were obtained and differentially expressed miRNAs were identified by the criteria of p-values less than 0.01 and fold-changes greater than ±2-fold. As a criterion for reliably expressed miRNAs, we further screened differentially expressed miRNA level for those with mean counts greater than the global mean value. In Figure 16 A, the volcano plot represents the p-values as a function of fold-changes is shown. By these criteria 16 miRNAs were differentially expressed as indicated by the data points with red circles.

Technical Replicates of microRNA profiles. Several samples from urine, serum, or lavage were analyzed multiple times which provided an opportunity to compare the similarity of technical replicates of the reproducibility of the RNA isolation and the NanoString profiles. Two replicate profiles are compared in **Figure 7** for the analysis of urine from STAMPEDE subject 1 from separate RNA isolations. The data reduction workflow included normalization after calculation of a global mean from all of the urine profiles. To correct for background, both assay values were taken as 50 counts or higher. Thirty-four miRNAs constituted the profiles for these samples and the numerical data is given in **Table 1**. Technical replicates were also calculated for miRNAs profiled in serum from subject 2 and these are displayed in **Figure 8**. **Table 2** lists the miRNAs that were expressed in two experiments and filtered as described for Table 1. Thirty-one of forty-six miRNAs that were above background and were expressed in both experiments showed standard deviations that were no more than 25% of the mean value. Technical replicates were also calculated for miRNAs profiled in lavage fluid from subject 2 and these are displayed in **Figure 9**. **Table 3** lists the miRNAs that were expressed in two experiments and filtered as described above. Fifteen of the thirty miRNAs that were above background and were expressed in both experiments showed standard deviations that were no more than 25% of the mean value.

Results

Proteomics analysis of lung fluid. Lung fluid was obtained by performing bronchoalveolar lavage on 15 control and 47 STAMPEDE subjects with dyspnea. Proteins were extracted and prepared for analysis by LC-MS(/MS). An AMT tag strategy was used to analyze the datasets produced by the mass spectrometer. An RMD-PAVS analysis (Matzke et al., 2011) identified outliers, 2 controls and 7 disease samples, which were removed from the analysis. The analysis identified 12,340 unique peptides corresponding to 987 proteins. A SPANS analysis (Webb-Robertson et al., 2011) was used to determine the optimal normalization for the peptide

abundance values. The optimal normalization was mean-centered, using rank invariant selected peptides with a p-value > 0.10. An MSstats analysis was performed removing proteins with insufficient observations to perform the hypothesis test. The remove of proteins with insufficient observations resulted in 652 proteins, which 79 (~12%) showed a significant difference in abundance (*p*-value < 0.05) between control and disease (66 proteins significantly greater and 13 significantly lower in disease compared to control). The 79 proteins were further analyzed by unsupervised hierarchical cluster analysis across all informative study subjects and the results are given in **Figure 3**. Close inspection of the dendrograms that group the samples revealed that give groups of subjects emerged with closely related patterns of protein expression. These patterns are visualized in more detail in **Supplementary Figures 1 – 6.** The subject groups, defined by differential expression of proteins are being compared for possible matches with the existing clinical diagnoses that were established by the STAMPEDE project. Results of this comparison, now in progress, will be reported elsewhere.

Proteomic analysis of urine samples. Proteins were extracted from urine obtained from 15 controls and 48 disease individuals and analyzed by LC-MS(/MS) using an AMT tag-based strategy. An RMD-PAVS analysis (6) identified outliers, 3 samples from the disease group, which were removed from the analysis, leaving a total of 45 in the disease group. The analysis identified 9,330 unique peptides corresponding to 846 proteins. A SPANS analysis (2) was used to determine the optimal normalization for the peptide abundance values. The optimal normalization was median-centered, using the top *L* Order Statistics peptide selection, with *L* being 614. An MSstats analysis was performed removing proteins with insufficient observations to perform the hypothesis test. The remove of proteins with insufficient observations resulted in 695 proteins, which 74 (~11%) showed a significant difference in abundance (*p*-value < 0.05) between control and disease (57 proteins significantly greater and 17 significantly lower in disease compared to control). Differentially expressed proteins derived from urine are presented

in **Figure 4** and compared with lavage results in **Figure 5.** While the control lavage profiles tended to cluster together (Figure 3) the control urine protein profiles were often flanked with profiles from subjects with dyspnea suggesting that the differences between the profiles were smaller. Nonetheless, groups of differentially expressed protein profiles may be recognized in the data from urine and these will be compared with existing diagnoses of the STAMPEDE subjects.

MicroRNA analysis of lung fluid

Bronchoalveolar lavage samples were profiled from 47 STAMPEDE subjects with dyspnea and 15 control individuals with no known lung abnormalities. While the NanoString profiling system can quantitate over 800 different miRNAs, only about 50 of these were routinely detected in typical samples in this study. The twenty-seven miRNAs that were most frequently observed in dyspnea or control profiles were listed in **Table 4**. MiRNAs 1246, 1283, and 630 were found in all 62 samples (dyspnea and control) in this study, while 24 miRNAs were detected in at least 56 of 62 (90%) of the samples profiled. The levels of many miRNAs such as 4443, 143-3p, 574-5p, and 378e were unchanged between dyspnea samples and controls. These may represent miRNAs that are usually expressed in the upper airways and would be sampled by a typical bronchial lavage. The levels of miRNAs 630, 575, and 489 in dyspnea samples were on average more than two-fold higher than the control average, while the level of miRNA 4516 in dyspnea samples was only half of the control average.

MiRNA signature groups in soldiers with dyspnea.

Groups of patients were recognized by inspection of the hierarchical cluster results that tended to be grouped together because they showed similar subsets of miRNAs at similar levels. Group 1, for example, is visible at the bottom of **Figure 11** and in more detail in **Figure 12**. Patients 12-14, 17, 18 and 23 composed the basic group, although patients 11, 15, 16, 21, and 27 displayed

a related profile. This basic group of patients is defined by up-regulation of miRNAs 489, 187-3p, 212-3p, 1915-3p, 4488, and 4532. For each of these miRNAs group 1 expression is higher than expression in the controls ($p< 0.02$) or in the 37 other patients in the study ($p<0.03$) as shown in **Table 12**. Patients18, 14, 3, 17, 23 (but not patient 12) showed a pronounced decline in miRNA 21-5p which may be of mechanistic importance. It is possible that the 6 patients that define group 1 share a lung disease or have a diagnosis in common, that is also shared in some respects with patients 11, 15, 16, 21, and 27.

Group 2, also visible on Figure 12, was recognized in subjects 28, 35, 36, 42, 44, 45, 46 as a core group and patients 3, 5, 29, 41, 9, 8, 10, and possibly control 11. These subjects showed elevated expression of miRNAs 150-5p, 223-3p, 29b-3p, 200c-3p, let7g-5p, 342-3p, 15a-5p, 26b-5p, 142-3p 16-5p, 343a-5p, 93-5p, 191-5p. More complete statistical analysis for this group is in progress.

Groups 3, 4 and 5 were defined from microRNAs that were expressed by most dyspnea and control subjects, but were very highly expressed in lavage samples from certain subjects, usually with dyspnea, but not others. The expression range for the defining miRNAs varied from 66-fold to over 100-fold (**Table 7**). Group 3 was defined by 10 subjects (9 with dyspnea and one control) with elevated expression of miRNA 320e and ten subjects with low expression as listed in **Table 6**. The normalized molecular counts for this miRNA were plotted from highest to lowest as shown in **Figure 13**. Group 4 was defined by ten subjects with elevated expression of miRNA 630 as shown in **Figure 14**. Only subjects with dyspnea expressed miRNA 630 at the highest levels as shown in Table 6. Group 5 was defined by ten subjects (control and dyspnea) with elevated expression of miRNA 4516 as shown in Table 6 and **Figure 15.** The mean high and low expression values for these miRNAs are presented in Table 7 which shows that the p-values that distinguish the high from the low expression groups were 10^{-5} or lower. Conceivably, these three groups, each defined by a single miRNA may be indicators of a process in the lung

that might occur occasionally in anyone, but was active, chronic, or exaggerated in STAMPEDE subjects and a few controls at the time the lavage was collected. Possible processes could include inflammation, low-level fibrosis, an atopic reaction, low-level infection, or excess mucus production.

Group 6 and 7 were derived by a slightly different data reduction workflow included an explicit false discovery rate (< 0.01) and more stringent expression thresholds (greater or less than 2-fold) is pictured in **Figure 16**. Group 6 includes dyspnea subjects 16, 18, 13, 17, 15, 21 (and possibly 12, 23, 14 and 11) and is defined by elevated expression of miRNAs 371a-5p, 187-3p, 1915-3p, 4488, and 421 relative to controls and other subjects but decreased expression of miRNAs125b-5p, let 71-5p, 191-5p and 631 as shown in Fig. 16 C (upper left). Thus group 6 overlaps with the STAMPEDE subjects in group 1 since both groups share up-regulated expression of miRNAs 187-3p and 4488. Group 7 includes dyspnea subjects 46, 34, 45, 44, 9, 28, 36 (and possibly 39, 37, 22, 35, & 43) and it is defined by elevated expression of miRNAs 191-5p, let-7i-5p, and 125b-5p but lower expression of miRNAs 371a-5p, 187-3p, 1915-3p, 4488, 421, 663a relative to other subjects with dyspnea and the controls. The STAMPEDE subjects that were placed in group 6 and 7 were summarized in **Table 8.** Indeed, the up- and down regulated miRNAs of group 6 appear to be reversed in group 7. One practical consequence of this is that there are many top-scoring pairs (i.e., 371a-5p and 125b-5p) that alone could distinguish a patient in group 6 from a patient in group 7 or from the typical control individual. Clearly, there are many such pairs of miRNAs with reciprocal expression. Calculation of top scoring pairs and other statistical tests that distinguish group 6 from group 7 from the controls are in progress.

MicroRNA analysis of urine and serum.

Complete profiles of miRNAs were obtained for all the urine samples and nearly all of the serum samples that were provided from the 48 STAMPEDE subjects and the 15 control individuals. Data analysis is still in progress and it will be reported elsewhere.

Discussion

In this study we applied advanced protein and RNA profiling methods to identify potential molecular markers that correlate with lung diseases that may be present in the active duty soldiers with dyspnea that enrolled in the STAMPEDE project. While early studies suggested that overseas deployment to Iraq or Afghanistan was associated with an increased risk of asthma (Szema et al 2010) or constrictive bronchiolitis (King et al., 2011) the number of soldiers who self-reported with dyspnea remained low. Nonetheless, they may be at risk for these or many other lung disorders. In part this was the rationale for the creation of the STAMPEDE project: to evaluate as many soldiers with dyspnea as possible at one location, develop diagnoses with conventional clinical tests, and retain the patient registry for possible follow-up. Samples of urine, serum and bronchoalveolar lavage were collected from a first cohort of STAMPEDE subjects. Urine and lavage samples were profiled for protein while urine, lavage, and serum were profiled for miRNAs. Since provisional diagnoses have been established for the STAMPEDE subjects, their diagnoses can now be compared to the groups of differentially expressed proteins or the miRNAs that these subjects expressed. If one or more of the molecular profiles matches study subjects with the same conventional diagnoses, the molecular profiles become candidate biomarkers for that diagnosis. This would make it possible to supplement conventional lung disease diagnosis with a molecular profile. Such molecular profiles could become clinically useful biomarker profiles after additional validation studies.

The five protein groups and seven miRNA groups are now being compared with the diagnoses for the STAMPEDE subjects. At least some of the STAMPEDE subjects who were placed in the miRNA group 1 (and by extension, group 6) had been diagnosed with asthma. Now that molecular groups have been defined, all of the relevant clinical data such as cell counts and cytokine levels from the lavage fluid are being compiled. When these comparisons are complete, all of the results will be published in the regular literature.

The data from the controls is valuable because for the first time it establishes 'wellness' for normal or typical individuals, not known to have active lung disorders. Biomarkers of wellness may of themselves be valuable in the future to judge lung health in routine physicals, return to normalcy after a lung procedure or disease, or a response to drug therapy for a lung condition such as asthma, fibrosis, or cancer.

Another unexpected result was the finding that several miRNAs were expressed by most subjects with or without dyspnea. While some were expressed at about the same level in all subjects, others were expressed at quite different levels among study subjects and these became the basis for study groups 3, 4, 5. We speculate that these may be derived from a fundamental lung cell or tissue such as bronchial smooth muscle or alveolar epithelium or alternatively from a cell that enters the lung from the circulation such as a macrophage, lymphocyte, or eosinophil. Departures from low- or baseline expression could be an indication of a disease or some other pathologic process. We are also investigating whether any of the differentially expressed miRNAs could be targeting the mRNAs for some of the differentially expressed proteins that were observed in this study.

Acknowledgements. We thank David Jackson (USACEHR) for his encouragement and support during the course of this work as well as our colleagues at PNNL and Kai Wang, David Huang, and Sara McClarty, at the ISB.

Key Research Accomplishments

1. Differentially expressed proteins after dust instillation in the pre-clinical rat model.
2. Differentially expressed proteins after silica instillation in the pre-clinical rat model.
3. Differentially expressed miRNAs after dust instillation in the pre-clinical rat model.
4. Differentially expressed miRNAs after silica instillation in the pre-clinical rat model.
5. Five groups of differentially expressed proteins in brochoalveolar lavage fluid that may correlate with and thus be biomarkers for discrete lung disorders in soldiers with dyspnea.
6. Expression levels of 79 proteins from brochoalveolar lavage fluid from normal individuals that define wellness or a healthy lung.
7. Several groups of differentially expressed proteins in urine that may correlate with and thus be biomarkers for discrete lung disorders in soldiers with dyspnea.
8. Expression levels of 74 proteins from urine from normal individuals not known to have a lung disorder that may also define wellness or a healthy lung.
9. Seven groups of differentially expressed miRNAs in bronchoalveolar lavage fluid that may correlate with and thus be biomarkers for discrete lung disorders in soldiers with dyspnea.
10. The expression pattern of a group of miRNAs from bronchoalveolar lavage fluid from individuals with healthy lungs that define lung health or wellness.
11. Groups of differentially expressed miRNAs in urine are being defined and these will be tested for correlation with conventional lung disease diagnoses in soldiers with dyspnea.
12. Groups of differentially expressed miRNAs in serum are being defined. These groups will be tested for correlation with conventional lung disease diagnoses in soldiers with dyspnea.

Reportable Outcomes

We plan to report the detailed differentially expressed protein data and miRNA data from the pre-clinical rat dust instillation study. The findings from silica-treated animals were consistent

with a fibrotic response and this has not been described well in the regular literature with the detail we can provide.

We are preparing to publish the protein and miRNA profiling results from the lavage and urine studies in collaboration with Dr. Michael Morris. Not only are one or more protein or miRNA groups likely to match with one or more conventionally diagnosed lung disorders, the patterns from the control sample donors define lung health or wellness in extraordinary detail.

Thanks to the extension of our contract, the collaboration that was enabled with Dr. Michael Morris and his colleagues at the SAMMC has already introduced the military health care system to the results and the promise of this research program.

Two papers are in preparation that will summarize the results of this Contract.

- Gelinas R, Wang K, Brown J. et al. 2013. Protein and miRNA profiling of lavage fluid from a pre-clinical dust-instillation model in rats. (in preparation).
- Brown J, Morris MJ, & Gelinas R et al., 2013. Protein and miRNA profiles from bronchoalveolar lavage or urine associated with diagnosed lung disease from soldiers with dyspnea. (in preparation).

List of personnel supported by this Contract at ISB.

- Richard Gelinas, Senior Scientist
- Kai Wang, Senior Scientist

Conclusions

Profiling of proteins and miRNAs using advanced methods can give insights into the most detailed pathological as well as normal physiologic processes. The protein and miRNA profiles we described for the dust-instillation model in rats may be useful in defining acute processes

such as inflammation or more chronic processes such as fibrosis, after validation and confirmation with human samples. The marker groups we have identified in soldiers with dyspnea may be closely related to lung disorders such as asthma or bronchiolitis. As these correlations are made the candidate markers we described would be ready for translation into clinical trials. The ultimate outcome would be novel platforms for new objective information to speed the reliable diagnosis of lung disorders.

References

1. Kelly, R. T., Page, J. S., Luo, Q., Moore, R. J., Orton, D. J., Tang, K., and Smith, R. D. (2006) Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry. Anal Chem 78, 7796-7801

2. Webb-Robertson, B. J., Matzke, M. M., Jacobs, J. M., Pounds, J. G., and Waters, K. M. (2011) A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through dataset-dependent ranking of normalization scaling factors. Proteomics 11, 4736-4741

3. Clough, T., Key, M., Ott, I., Ragg, S., Schadow, G., and Vitek, O. (2009) Protein quantification in label-free LC-MS experiments. J Proteome Res 8, 5275-5284

4. Benjamini, Y., and Hochberg, Y. (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met 57, 289-300.

5. Team, R. D. C. (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

6. Matzke MM, Waters KM, Metz, T.O., Jacobs, J.M. Sim, A.C., Bric, R.S., Pounds, J.G., and Webb-Robertson, B.J. (2011) Improved quality control processing of peptide-centric LC-MS proteomics data. Bioinformatics 27, 2866-2872.

7. Maiolica A, Borsotti D, Rappsilber J. 2005. Self-made frits for nanoscale columns in proteomics. *Proteomics* **5**, 3847-3850).

8. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*. 34:374-378.

9. Waggott et al. 2012. NanoStringNorm . *Bioinformatics*. **28**: 1546-1548.

10. Szema AM, Peters MC, Weissinger KM, Gagliano CA, Chen JJ. 2010. New-onset asthma among soldiers serving in Iraq and Afghanistan. *Allergy Asthma Proc.* **31**: 67-71. doi: 10.2500/aap.2010.31.3383.

11. King MS, Eisenberg R, Newman J, Tolle J, et al. 2011. Constrictive bronchiolitis in soldiers returing from Iraq and Afghanistan. N Engl J. Med **365**: 222-230. doi: 10.1056/NEJMoa1101388.

Supporting Tables and Figures

Table 1 : Technical replicates for miRNA profiles for urine from one subject.

MicroRNA	Exp.1	Exp. 2	mean	std dev	as % mean
miR-21-5p	415	273	344	100.2	29
miR-23c	104	83	94	14.8	16
miR-25-3p	70	170	120	70.5	59
miR-95	51	71	61	13.8	23
miR-125b-5p	155	174	164	13.5	8
miR-143-3p	119	123	121	2.6	2
miR-144-3p	99	118	108	13.4	12
miR-155-5p	78	110	94	22.9	24
miR-199a-5p	52	54	53	1.8	3
miR-200a-3p	65	83	74	13.0	18
miR-212-3p	69	138	103	48.5	47
miR-222-3p	72	69	70	2.1	3
miR-302d-3p	70	154	112	59.5	53
miR-320e	60	87	74	19.0	26
miR-363-3p	173	143	158	21.4	14
miR-378e	263	346	305	58.5	19
miR-495	84	98	91	9.6	11
miR-504	52	80	66	19.7	30
miR-518b	67	51	59	11.5	20
miR-548ai	48	72	60	17.0	28
miR-598	235	177	206	40.6	20
miR-631	55	94	75	27.5	37
miR-663b	84	71	77	9.6	12
miR-761	104	92	98	8.0	8
miR-769-5p	64	58	61	4.4	7
miR-1183	206	179	193	18.8	10
miR-1253	73	89	81	11.0	14
miR-1246	94	120	107	18.1	17
miR-1273d	73	87	80	10.2	13
miR-1283	631	636	633	3.0	0.5
miR-1286	55	83	69	19.8	29
miR-1827	56	94	75	27.0	36
miR-4443	244	168	206	53.3	26
miR-4516	169	989	579	579.5	100

Table 2 : Technical replicates for miRNA profiles from serum from one subject.

miRNA	expt. 1	expt. 2	mean	st dev	st dev % mean
let-7g-5p	52	79	65	18.82301	28.75
miR-16-5p	60	239	150	126.9688	84.83
miR-21-5p	303	512	408	147.6209	36.21
miR-23c	174	87	130	61.3724	47.04
miR-25-3p	133	256	195	86.78761	44.58
miR-107	51	52	52	0.448871	0.87
miR-125b-5p	187	259	223	50.738	22.73
miR-141-3p	70	59	64	7.397532	11.50
miR-143-3p	109	174	142	45.95844	32.45
miR-144-3p	118	146	132	20.02125	15.17
miR-155-5p	98	118	108	14.57433	13.51
miR-188-5p	65	78	71	9.21092	12.93
miR-199a-5p	79	61	70	12.76169	18.18
miR-200a-3p	79	73	76	4.699051	6.19
miR-211-5p	51	52	52	0.448871	0.87
miR-222-3p	98	92	95	3.74985	3.95
miR-302d-3p	88	134	111	32.39892	29.24
miR-320e	82	171	127	62.88379	49.68
miR-363-3p	200	126	163	51.9609	31.84
miR-371a-3p	52	56	54	2.69773	4.99
miR-378e	217	270	243	36.89439	15.16
miR-451a	381	1359	870	691.8142	79.54
miR-489	75	58	67	12.22855	18.34
miR-495	125	90	107	24.34004	22.66
miR-504	68	77	72	6.428926	8.91
miR-542-3p	51	52	52	0.448871	0.87
miR-548ai	58	60	59	1.531544	2.59
miR-548z	58	60	59	1.531544	2.59
miR-556-5p	57	53	55	2.916216	5.31
miR-570-3p	58	80	69	15.45792	22.44
miR-598	259	152	206	75.30028	36.62
miR-630	62	57	59	3.399394	5.72
miR-631	82	82	82	0.151375	0.18
miR-663b	71	51	61	14.62729	23.93
miR-761	126	82	104	30.88678	29.78
miR-766-3p	59	58	58	0.617399	1.06
miR-769-5p	94	61	77	23.00683	29.71
miR-1183	254	211	233	30.10611	12.94
miR-1246	94	159	126	45.89208	36.38

miRNA	expt. 1	expt. 2	mean	st dev	st dev % mean
miR-1253	78	106	92	19.4389	21.13
miR-1273d	91	66	79	17.29296	22.01
miR-1277-3p	52	51	51	0.967105	1.88
miR-1283	558	763	661	144.6845	21.90
miR-128	77	75	76	1.867099	2.46
miR-1827	65	62	63	1.783586	2.81
miR-4443	302	221	262	57.65986	22.04

Table 3. Technical replicates of miRNA profiles for lavage from one subject. (Ref: 0408stamp1-2,5-10.xlsx)

miRNA	expt.1	expt. 2	mean	st dev	st dev % mean
miR-21-5p	385	210	297	123.9	42
miR-23c	97	51	74	32.2	44
miR-25-3p	266	187	227	55.9	25
miR-125b-5p	140	172	156	22.7	15
miR-143-3p	183	91	137	65.1	48
miR-144-3p	68	76	72	5.9	8
miR-155-5p	74	58	66	11.9	18
miR-222-3p	72	52	62	14.1	23
miR-302d-3p	107	156	131	34.3	26
miR-363-3p	145	59	102	60.6	59
miR-378e	232	155	194	54.9	28
miR-495	124	66	95	40.7	43
miR-504	76	52	64	16.9	26
miR-514b-5p	69	53	61	11.1	18
miR-570-3p	80	78	79	1.3	2
miR-574-5p	98	70	84	20.0	24
miR-598	178	62	120	82.0	69
miR-612	68	76	72	5.9	8
miR-630	2818	159	1489	1880.6	126
miR-631	84	71	77	9.2	12
miR-720	272	249	260	16.1	6
miR-761	108	53	81	38.8	48
miR-1183	144	52	98	64.9	66
miR-1246	117	85	101	22.9	23
miR-1253	78	88	83	7.1	8
miR-1283	804	566	685	168.1	25
miR-1827	102	80	91	15.3	17
miR-4443	278	66	172	149.6	87
miR-4454	1431	704	1068	513.8	48
miR-4516	2133	75	1104	1455.0	132

Table 4. Frequently expressed miRNAs in lavage fluid.

miRNA	Number of pos samples (%)	Average of patient	Average of control	Stdev of patient	Stdev of control	Patient/ control
1246	62 (100)	204.8	243.2	14.6	144.2	0.8
1283	62 (100)	738.0	809.8	14.1	311.3	0.9
630	62 (100)	2719.4	1026.8	625.6	590.4	2.6
4516	61 (98)	4537.7	8310.6	37.8	6731.2	0.5
21-5p	61 (98)	946.7	1130.3	111.7	340.2	0.8
222-3p	61 (98)	126.2	134.1	11.3	37.0	0.9
25-3p	61 (98)	223.7	202.1	13.2	92.6	1.1
601	61 (98)	172.6	92.6	190.3	30.8	1.9
378e	60 (97)	321.0	310.7	18.5	150.5	1.0
574-5p	60 (97)	103.2	99.4	20.9	36.5	1.0
320e	60 (97)	3942.6	3182.3	84.6	2620.1	1.2
1183	59 (95)	184.7	237.6	12.5	98.4	0.8
598	59 (95)	204.3	249.7	18.0	109.2	0.8
363	59 (95)	188.8	224.9	38.3	118.2	0.8
143-3p	59 (95)	184.2	179.6	11.9	75.4	1.0
302d	59 (95)	156.3	137.0	19.4	51.3	1.1
4443	59 (95)	322.9	324.4	9.5	196.4	1.0
200a-3p	58 (94)	134.5	157.9	16.3	57.7	0.9
4454	58 (94)	1130.4	882.0	26.1	651.6	1.3
495	57 (92)	135.0	144.4	10.2	72.6	0.9
141-3p	56 (90)	93.8	111.6	9.5	27.8	0.8
761	56 (90)	108.5	123.1	16.3	48.3	0.9
570	56 (90)	119.2	91.9	46.3	34.7	1.3
489	56 (90)	142.0	67.4	2592.5	10.6	2.1
1286	55 (89)	91.5	84.1	15.0	25.1	1.1
575	55 (89)	183.5	78.2	204.8	30.8	2.3
1827	55 (89)	96.1	108.3	34.2	49.2	0.9

Table 5. MiRNAs that define group 1 of dyspnea subjects.

miRNA	mean	st dev	P vs controls	P vs patients
489	495	195	0.0024	0.003
187-3p	1077	580	0.0068	0.0074
212-3p	751	561	0.0285	0.0305
1915-3p	531	200	0.0015	0.0016
4488	303	106	0.0014	0.00154
4532	822	401	0.0049	0.0051

Table 6. Study subjects in groups 3, 4, and 5.

MiRNA 320e: group 3		MiRNA 630: group 4		MiRNA 4516: group 5	
High expression	Low expression	High expression	Low expression	High expression	Low expression
stamp 30	stamp 11	stamp 44	stamp 47	control 6	stamp 24
stamp 7	stamp 45	stamp 28	stamp 14	control 4	stamp 1
stamp 6	stamp 25	stamp 10	stamp 11	stamp 7	stamp 46
stamp 32	stamp 15	stamp 43	stamp 13	control 7	stamp 34
stamp 4	stamp 20	stamp 22	stamp 15	stamp 37	stamp 45
stamp 37	stamp 46	stamp 35	stamp 12	control 13	stamp 9
control 2	stamp 26	stamp 6	stamp 23	stamp 38	stamp 25
stamp 31	control 10	stamp 32	stamp 25	stamp 13	stamp 42
stamp 19	control 14	stamp 36	stamp 5	control 5	control 14
stamp 39	stamp 5	stamp 34	control 14	stamp 39	stamp 5

Table 7. Expression levels of groups 3, 4, & 5.

Group	MiRNA	hi express, mean	st deviation	low express, mean	st deviation	p value
3	320e	13581.8	5524.0	195.7	118.0	6.15E-05
4	630	6780.1	179.9	1946.2	72.3	1.96E-06
5	4516	15233.5	264.9	1934.6	133.4	1.35E-09

Table 8. STAMPEDE subjects in groups 6 & 7.

	Group 6			Group 7	
1	STAMPEDE subject	16	1	STAMPEDE subject	46
2	STAMPEDE subject	18	2	STAMPEDE subject	34
3	STAMPEDE subject	13	3	STAMPEDE subject	45
4	STAMPEDE subject	17	4	STAMPEDE subject	44
5	STAMPEDE subject	15	5	STAMPEDE subject	9
6	STAMPEDE subject	21	6	STAMPEDE subject	28
7	STAMPEDE subject	12	7	STAMPEDE subject	36
8	STAMPEDE subject	23	8	STAMPEDE subject	39
9	STAMPEDE subject	14	9	STAMPEDE subject	37
10	STAMPEDE subject	11	10	STAMPEDE subject	22
			11	STAMPEDE subject	35
			12	STAMPEDE subject	43

Figure Legends

Figure 1. Summary of correlation scores between datasets representing individuals. Panels A, C, and E are of lung fluid proteomes and panels B, D, and F are urine proteomes. The 56 bars in panel A and the 60 bars in panel B represent the mean correlation for a dataset across the biological replicates. The red horizontal line in panels A and B indicates the mean correlation threshold used to distinguish outliers, for lung fluid and urine, respectively. Outlier datasets are indicated by a red bar within the plots, while controls are green and disease are purple. Panels C and D are correlation heatmaps prior to outliers being removed. The color of the cells in the heatmap correspond to the pairwise correlation coefficients between the row/column datasets, with red representing a perfect correlation (+1) and blue the minimal correlation value in the matrix. Panels E and F are the correlation heatmaps after outliers have been removed. The green and purple bars above and to the left of the correlation heatmaps designate control and disease, respectively.

Figure 2. PCA plot of BALF and urine datasets on left and right, respectively. Each dot represents an individual with green indicating control and purple designating disease individuals. Green and purple ellipses indicate the distribution of each group within the dimensions of the first and second principal components.

Figure 3. **A.** Heatmap of the 79 significantly different proteins in lung fluid between control and disease individuals, designated by the light and dark blue bars above the heatmap, respectively. The protein abundance values were scaled using z-score, with red representing 1 standard deviation above the mean and green being 1 standard deviation below the mean. Uniprot accession identifiers for the proteins are shown on the right side of the heatmap**. B**. The data of A, replotted with groups of subjects (control or dyspnea) that have closely related profiles of differenetially expressed proteins identified as groups 1 through 5 via the shading in the dendrogram (top of figure).

Figure 4. Heatmap of the 74 significantly different proteins in urine between control and disease individuals, designated by the light and dark blue bars above the heatmap, respectively. The protein abundance values were scaled using z-score, with red representing 1 standard deviation above the mean and green being 1 standard deviation below the mean. Uniprot accession identifiers for the proteins are shown on the right side of the heatmap.

Figure 5. Heatmap showing the 79 and 74 significantly different proteins in lung fluid (left) and urine (right). Proteins and subjects have been clustered by hierarchical clustering, with dendrograms on top and left of map depicting distance measured calculated as Pearson product-moment correlation coefficients. Protein abundances were scaled by z-score, with green and red representing 1 standard deviation below and above the mean, respectively. The bar above the heatmap indicates controls (green) and disease (purple) subjects.

Figure 6. Scatterplot of lung fluid (left) and urine (right) proteins. Significant proteins have an adjusted p-value less than 0.05. Proteins with significantly greater and lower abundance in disease are shown in red and green, respectively, with the actual number displayed at top of plot.

Figure 7. Analysis of technical replicates of miRNA levels from urine. MiRNA from three separate isolations was profiled by NanoString. The mean normalized counts are displayed along with one standard deviation. (ref: 0403stamp1-4.xlsx)

Figure 8. Analysis of technical replicates of miRNA levels from serum. MiRNA from two separate isolations was profiled by NanoString. The mean normalized counts are displayed along with one standard deviation. (ref: 0403stamp1-4.xlsx)

Figure 9. Analysis of technical replicates of miRNA levels from lavage. MiRNA from two separate isolations from subject 2 was profiled by NanoString. The mean normalized counts are displayed along with one standard deviation. (ref: 0408stamp1-2,5-10.xlsx)

Figure 10. Hierarchical cluster analysis of miRNA profiles from lavage fluid. MiRNA profiles from lavage samples from all dyspnea and control subjects were clustered (Pearson distance) and a portion of the data that includes about 70 miRNAs that were expressed at or above the lower detection limit (50 normalized molecular counts) are displayed. Note that the color assignments are non-linear where green corresponds to 50 moluecular counts, black is 500 counts, and red is greater than or equal to 10,000 counts (see scale bar at top). MiRNAs 320e, 4516, and 630 were expressed by most subjects and are labeled on the right. A group of six miRNAs that was elevated in a group of 5 STAMPEDE subjects relative to controls is highlighted at the bottom. This group of subjects and miRNAs constitutes 'group 1' from this study which is being reviewed by Dr. Michael Morris to determine if they have been diagnosed with the same lung disease.

Figure 11. Cluster analysis of selected miRNAs to highlight the miRNAs and dyspnea subjects in group 1. MiRNAs 489, 187-3p, 212-3p, 1915-3p, 4488, 4532 and possibly 371a-5p are differentially elevated in group 1 subjects relative to controls while miRNAs 320e, 21-5p, 630, and 4516 are widely expressed among controls as well as dyspnea subjects. Dyspnea subjects 12-14, 17, 18 and 23 compose the basic group, although patients 11, 15, 16, 21, and 27 display a related profile.

Figure 12. Cluster analysis of selected miRNAs to highlight the miRNAs dyspnea subjects in group 2. Dyspnea subjects 28, 35, 36, 42, 44, 45, 46 define group 2 and subjects 3, 5, 29, 41, 9, 8, 10, and possibly control 11 may be related. These subjects showed elevated expression of miRNAs 150-5p, 223-3p, 29b-3p, 200c-3p, let7g-5p, 342-3p, 15a-5p, 26b-5p, 142-3p 16-5p, 343a-5p, 93-5p, 191-5p.

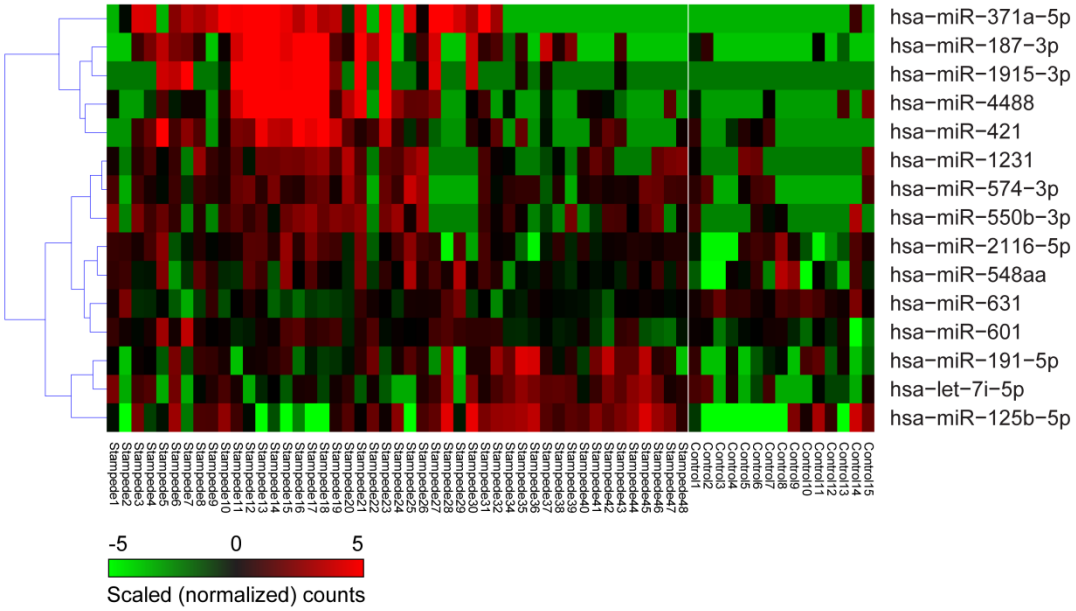
Figure 13. The expression of miRNA 320e in dyspnea subjects and controls plotted from highest (left) to lowest (right) along the x-axis. The ten subjects with highest expression were

defined as group 3 and compared with the ten subjects with lowest expression as a reference. The expression of miRNA 1283 which was detected in most samples but varied only slightly between individuals was plotted for reference.

Figure 14. The expression of miRNA 630 in dyspnea subjects and controls plotted from highest (left) to lowest (right) along the x-axis. The ten subjects with highest expression were defined as group 4 and compared with the ten subjects with lowest expression as a reference. The expression of miRNA 1283 which was detected in most samples but varied only slightly between individuals was plotted for reference.

Figure 15. The expression of miRNA 4516 in dyspnea subjects and controls plotted from highest (left) to lowest (right) along the x-axis. The ten subjects with highest expression were defined as group 5 and compared with the ten subjects with lowest expression as a reference. The expression of miRNA 1283 which was detected in most samples but varied only slightly between individuals was plotted for reference.

Figure 16. Derivation of dyspnea subject groups six and seven. **A.** Volcano plot in which the red circles indicate the miRNA that met the criteria of being detected with a P-value<0.01 and a (base ten) fold-change of >±2. **B.** Hierarchical cluster analysis of the differentially expressed miRNAs from A. To highlight differences between groups, the normalized counts of each miRNA were scaled to have a mean (log2) of zero across samples. Red and green represent higher and lower abundance, respectively. MiRNAs 371a-5p, 187-3p. 1915-3p. 4488, and 421 tend to be co-expressed in STAMPEDE subjects 10-20 (upper left) while elevations in the levels of miRNAs 191-5p, let-7i-5p and 125b-5p tend to occur in STAMPEDE subjects 28-36 & 41-46 (center) **C.** Hierarchical clustering of both differentially expressed miRs as well as subjects. Here, while most of control samples again clustered by themselves, STAMPEDE subjects again separated into two groups as before.



Supplementary Figure 1. Significant proteins in BALF. The dendrogram from Fig. 3B is shown indicating the study subjects (control or dyspnea) with distinct patterns of differential protein expression that define five groups.

Supplementary Figure 2. The study subjects and key proteins that define differential protein expression group 1.

Supplementary Figure 3. The study subjects and key proteins that define differential protein expression group 2.

Supplementary Figure 4. The study subjects and key proteins that define differential protein expression group 3.

Supplementary Figure 5. The study subjects and key proteins that define differential protein expression group 4.

Supplementary Figure 6. The study subjects and key proteins that define differential protein expression group 5.

Figure 1.

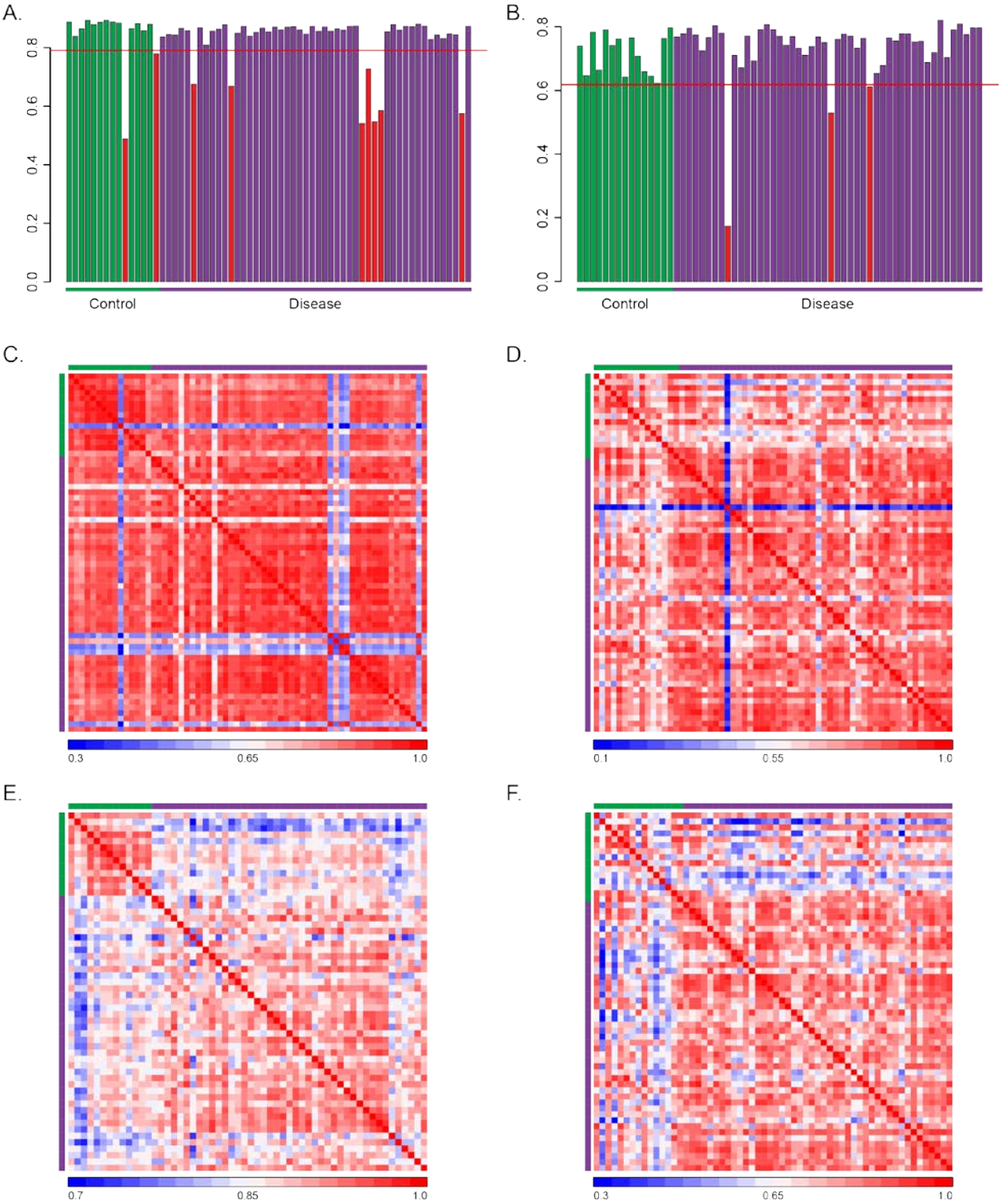

A.
0.0
0.2
0.4
0.6
0.8
Control
Disease
B.
0.0
0.2
0.4
0.6
0.8
Control
Disease
C.
0.3
0.65
1.0
D.
0.1
0.55
1.0
E.
0.7
0.85
1.0
F.
0.3
0.65
1.0

Figure 2.

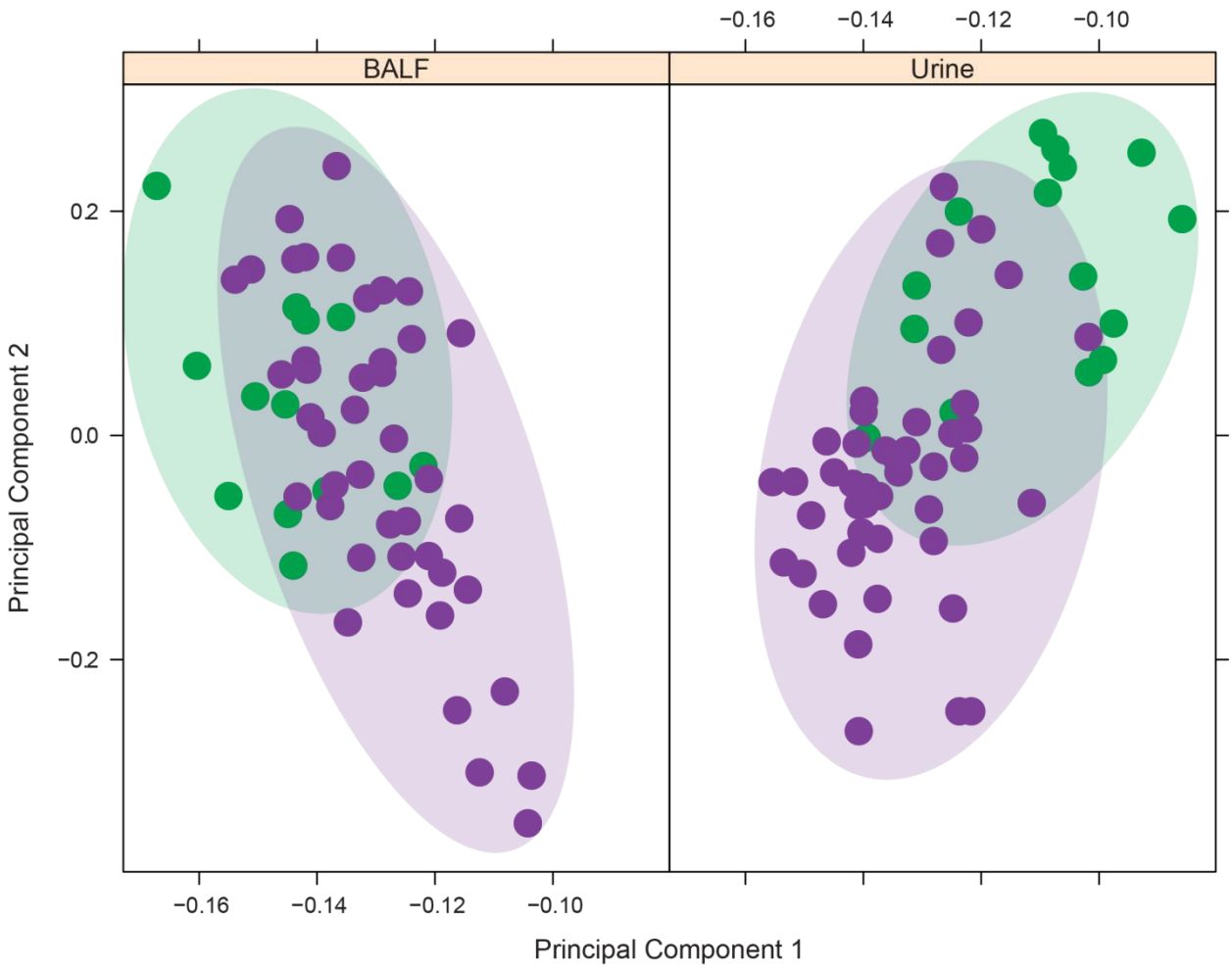

BALF
Urine
−0.16
−0.14
−0.12
−0.10
0.2
0.0
−0.2
Principal Component 2
Principal Component 1

Figure 3.

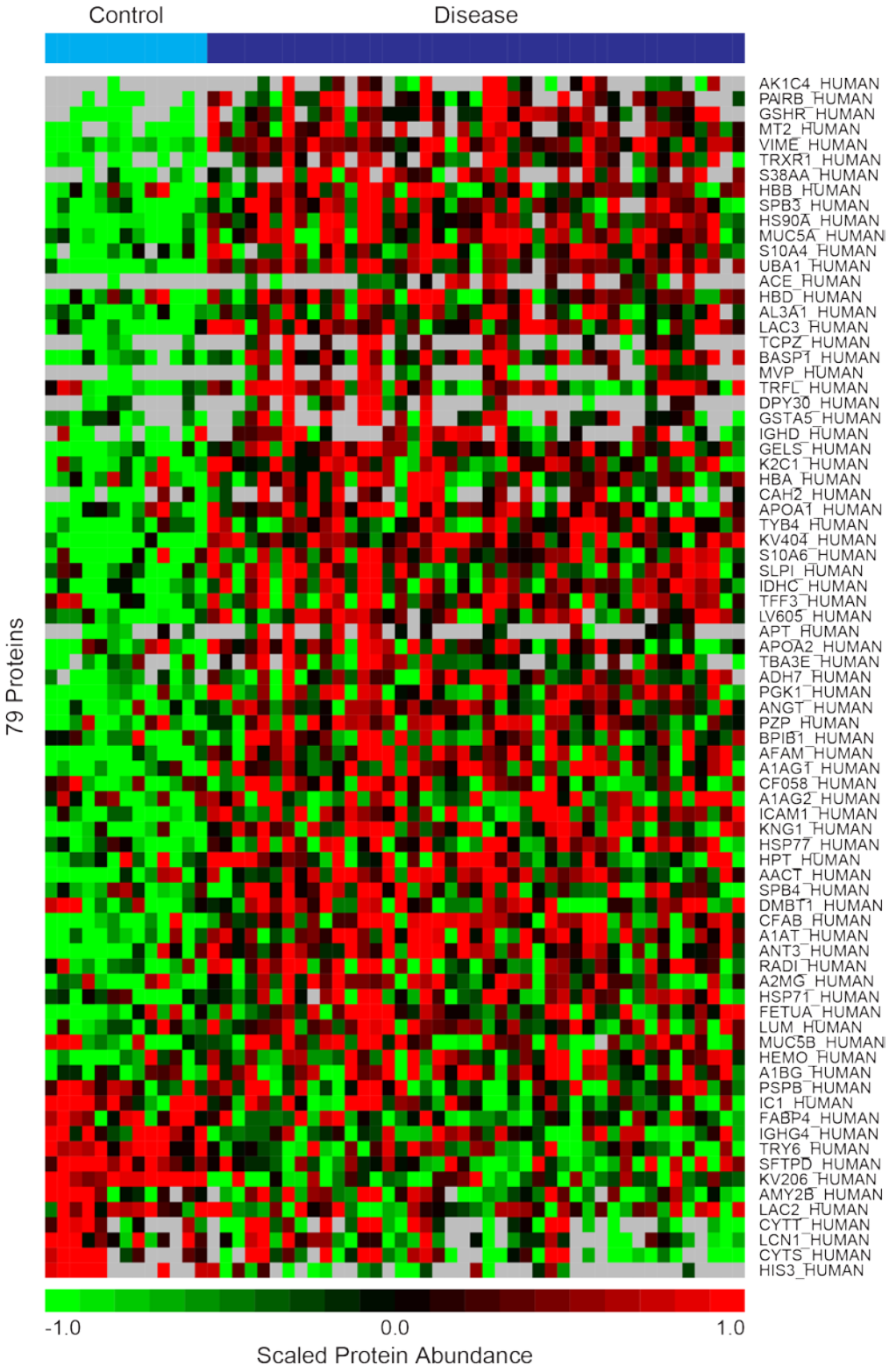

Control
Disease
79 Proteins
AK1C4_HUMAN
PARB_HUMAN
GSHR_HUMAN
MT2_HUMAN
VIME_HUMAN
TRXR1_HUMAN
S38AA_HUMAN
HBB_HUMAN
SPB3_HUMAN
HS90A_HUMAN
MUC5A_HUMAN
S10A4_HUMAN
UBA1_HUMAN
ACE_HUMAN
HBD_HUMAN
AL3A1_HUMAN
LAC3_HUMAN
TCPZ_HUMAN
BASP1_HUMAN
MVP_HUMAN
TRFL_HUMAN
DPY30_HUMAN
GSTA5_HUMAN
IGHD_HUMAN
GELS_HUMAN
K2C1_HUMAN
HBA_HUMAN
CAH2_HUMAN
APOA1_HUMAN
TYB4_HUMAN
KV404_HUMAN
S10A6_HUMAN
SLPI_HUMAN
IDHC_HUMAN
TFF3_HUMAN
LV605_HUMAN
APT_HUMAN
APOA2_HUMAN
TBA3E_HUMAN
ADH7_HUMAN
PGK1_HUMAN
ANGT_HUMAN
PZP_HUMAN
BPIB1_HUMAN
AFAM_HUMAN
A1AG1_HUMAN
CF058_HUMAN
A1AG2_HUMAN
ICAM1_HUMAN
KNG1_HUMAN
HSP77_HUMAN
HPT_HUMAN
AACT_HUMAN
SPB4_HUMAN
DMBT1_HUMAN
CFAB_HUMAN
A1AT_HUMAN
ANT3_HUMAN
RADI_HUMAN
A2MG_HUMAN
HSP71_HUMAN
FETUA_HUMAN
LUM_HUMAN
MUC5B_HUMAN
HEMO_HUMAN
A1BG_HUMAN
PSPB_HUMAN
IC1_HUMAN
FABP4_HUMAN
IGHG4_HUMAN
TRY6_HUMAN
SFTPD_HUMAN
KV206_HUMAN
AMY2B_HUMAN
LAC2_HUMAN
CYTT_HUMAN
LCN1_HUMAN
CYTS_HUMAN
HIS3_HUMAN
-1.0
0.0
1.0
Scaled Protein Abundance

Figure 3B.

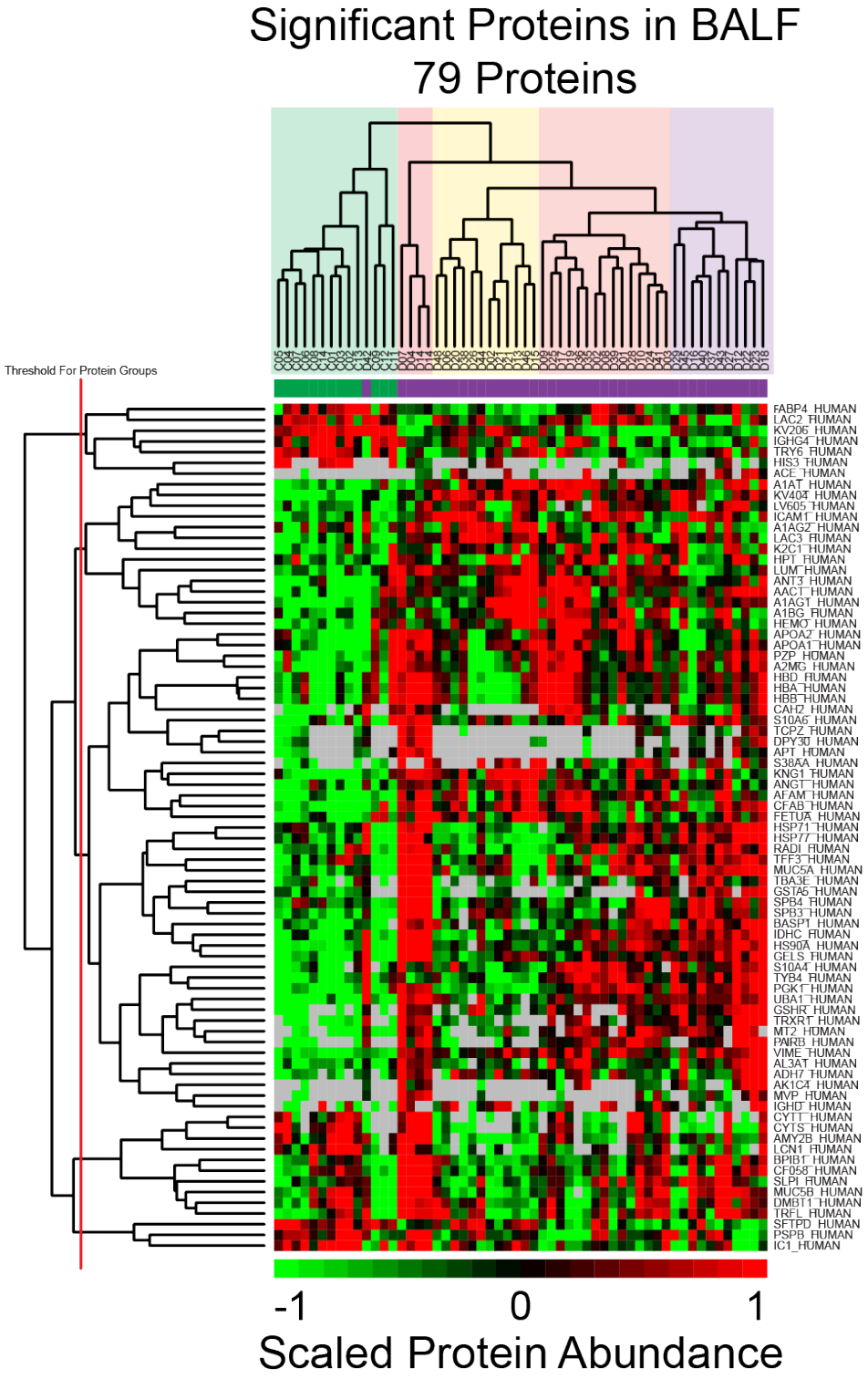

Significant Proteins in BALF
79 Proteins
Threshold For Protein Groups
FABP4_HUMAN
LAC2_HUMAN
KV206_HUMAN
IGHG4_HUMAN
TRY6_HUMAN
HIS3_HUMAN
ACE_HUMAN
A1AT_HUMAN
KV404_HUMAN
LV605_HUMAN
ICAM1_HUMAN
A1AG2_HUMAN
LAC3_HUMAN
K2C1_HUMAN
HPT_HUMAN
LUM_HUMAN
ANT3_HUMAN
AACT_HUMAN
A1AGT_HUMAN
A1BG_HUMAN
HEMO_HUMAN
APOA2_HUMAN
APOA1_HUMAN
PZP_HUMAN
A2MG_HUMAN
HBD_HUMAN
HBA_HUMAN
HBB_HUMAN
CAH2_HUMAN
S10A6_HUMAN
TCPZ_HUMAN
DPY30_HUMAN
APT_HUMAN
S38AA_HUMAN
KNG1_HUMAN
ANGT_HUMAN
AFAM_HUMAN
CFAB_HUMAN
FETUA_HUMAN
HSP71_HUMAN
HSP77_HUMAN
RADI_HUMAN
TFF3_HUMAN
MUC5A_HUMAN
TBA3E_HUMAN
GSTA5_HUMAN
SPB4_HUMAN
SPB3_HUMAN
BASP1_HUMAN
IDHC_HUMAN
HS90A_HUMAN
GELS_HUMAN
S10A4_HUMAN
TYB4_HUMAN
PGK1_HUMAN
UBA1_HUMAN
GSHR_HUMAN
TRXR1_HUMAN
MT2_HUMAN
PAIRB_HUMAN
VIME_HUMAN
AL3A1_HUMAN
ADH7_HUMAN
AK1C4_HUMAN
MVP_HUMAN
IGHD_HUMAN
CYTT_HUMAN
CYTS_HUMAN
AMY2B_HUMAN
LCN1_HUMAN
BPIB1_HUMAN
CF058_HUMAN
SLPI_HUMAN
MUC5B_HUMAN
DMBT1_HUMAN
TRFL_HUMAN
SFTPD_HUMAN
PSPB_HUMAN
IC1_HUMAN
-1
0
1
Scaled Protein Abundance

Figure 4.

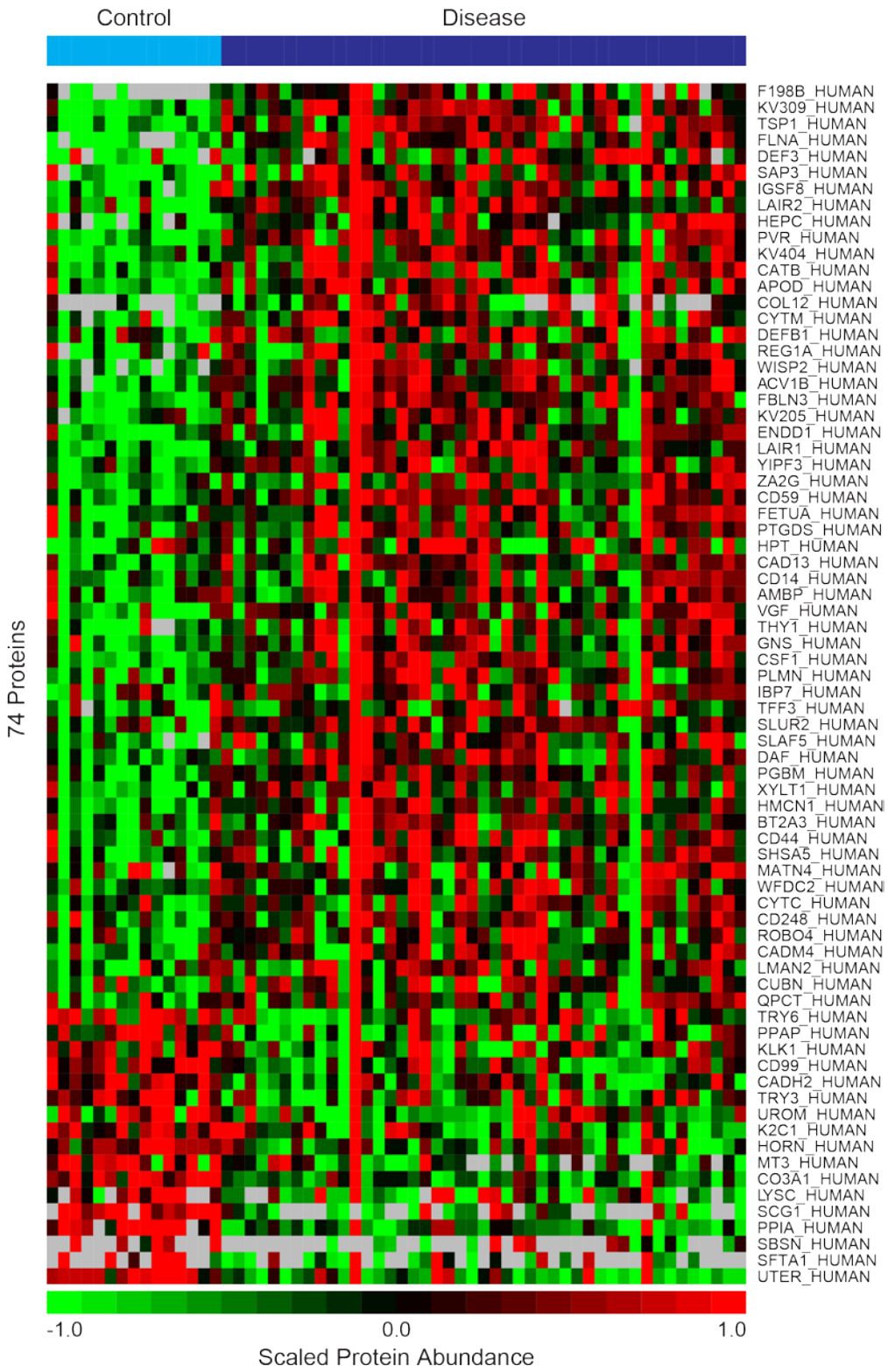

Control
Disease
74 Proteins
F198B_HUMAN
KV309_HUMAN
TSP1_HUMAN
FLNA_HUMAN
DEF3_HUMAN
SAP3_HUMAN
IGSF8_HUMAN
LAIR2_HUMAN
HEPC_HUMAN
PVR_HUMAN
KV404_HUMAN
CATB_HUMAN
APOD_HUMAN
COL12_HUMAN
CYTM_HUMAN
DEFB1_HUMAN
REG1A_HUMAN
WISP2_HUMAN
ACV1B_HUMAN
FBLN3_HUMAN
KV205_HUMAN
ENDD1_HUMAN
LAIR1_HUMAN
YIPF3_HUMAN
ZA2G_HUMAN
CD59_HUMAN
FETUA_HUMAN
PTGDS_HUMAN
HPT_HUMAN
CAD13_HUMAN
CD14_HUMAN
AMBP_HUMAN
VGF_HUMAN
THY1_HUMAN
GNS_HUMAN
CSF1_HUMAN
PLMN_HUMAN
IBP7_HUMAN
TFF3_HUMAN
SLUR2_HUMAN
SLAF5_HUMAN
DAF_HUMAN
PGBM_HUMAN
XYLT1_HUMAN
HMCN1_HUMAN
BT2A3_HUMAN
CD44_HUMAN
SHSA5_HUMAN
MATN4_HUMAN
WFDC2_HUMAN
CYTC_HUMAN
CD248_HUMAN
ROBO4_HUMAN
CADM4_HUMAN
LMAN2_HUMAN
CUBN_HUMAN
QPCT_HUMAN
TRY6_HUMAN
PPAP_HUMAN
KLK1_HUMAN
CD99_HUMAN
CADH2_HUMAN
TRY3_HUMAN
UROM_HUMAN
K2C1_HUMAN
HORN_HUMAN
MT3_HUMAN
CO3A1_HUMAN
LYSC_HUMAN
SCG1_HUMAN
PPIA_HUMAN
SBSN_HUMAN
SFTA1_HUMAN
UTER_HUMAN
-1.0
0.0
1.0
Scaled Protein Abundance

Figure 5.

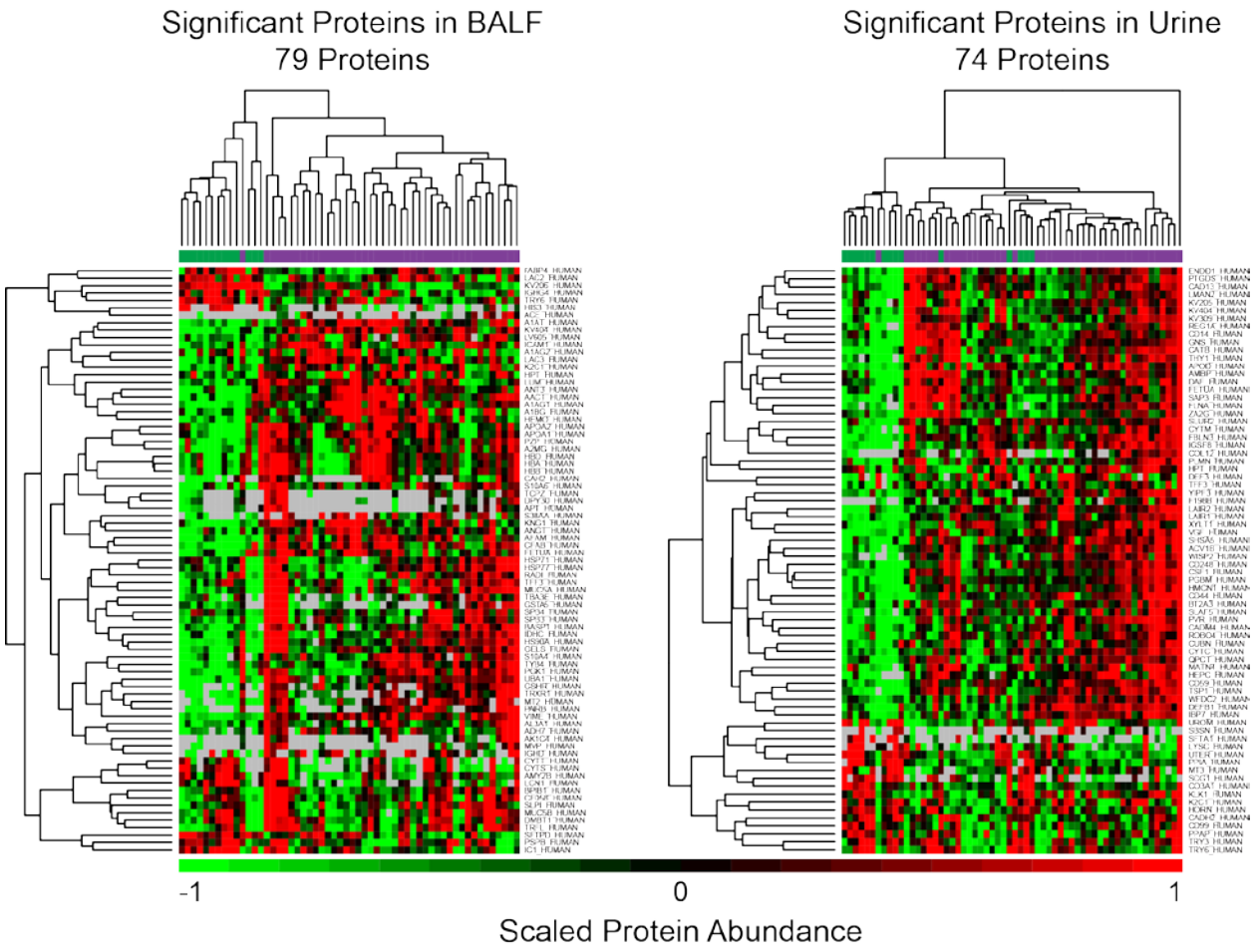

Significant Proteins in BALF
79 Proteins
Significant Proteins in Urine
74 Proteins
-1
0
1
Scaled Protein Abundance

Figure 6.

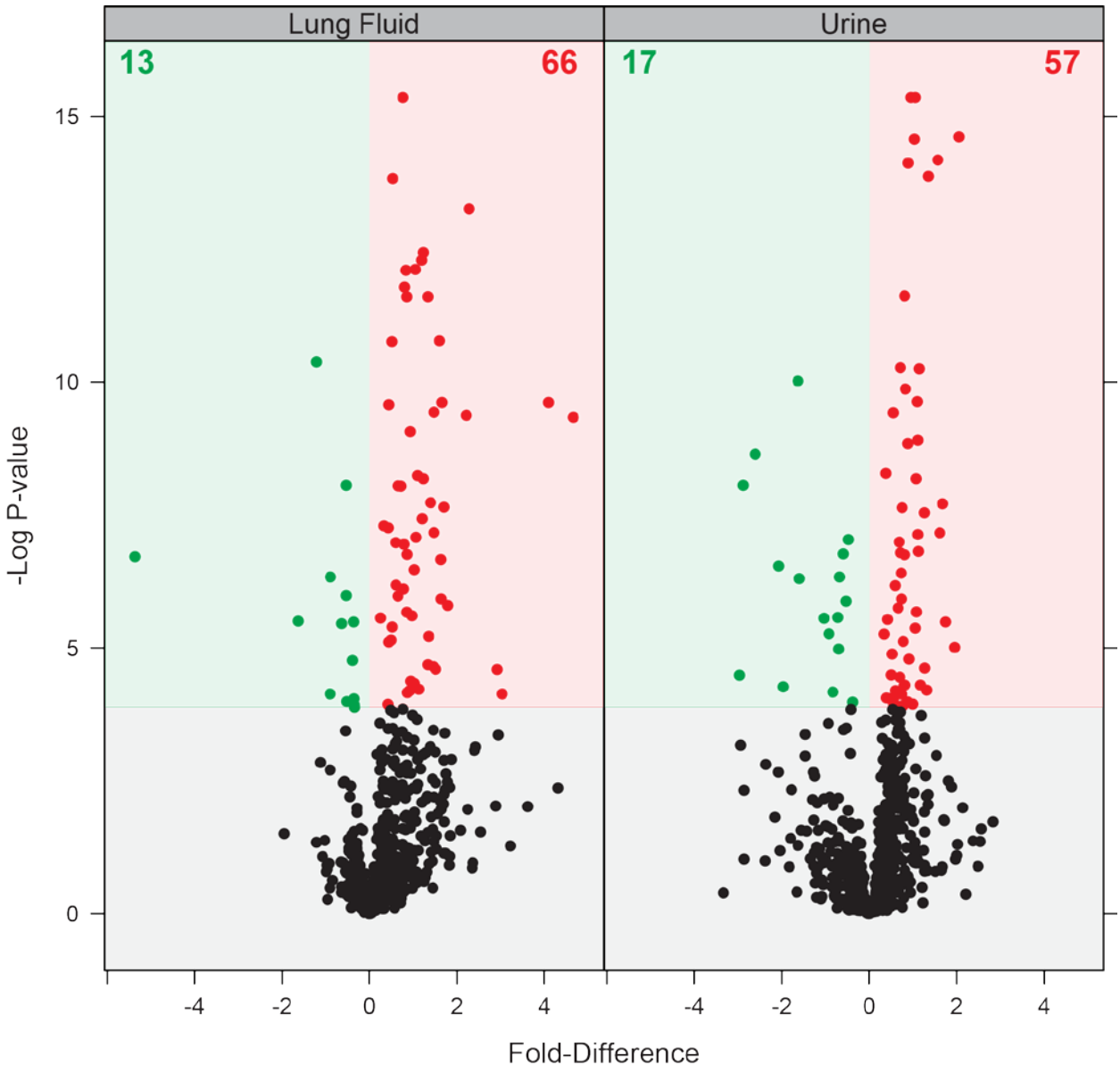

Lung Fluid
Urine
13
66
17
57
15
10
5
0
-Log P-value
-4
-2
0
2
4
Fold-Difference

Figure 7.

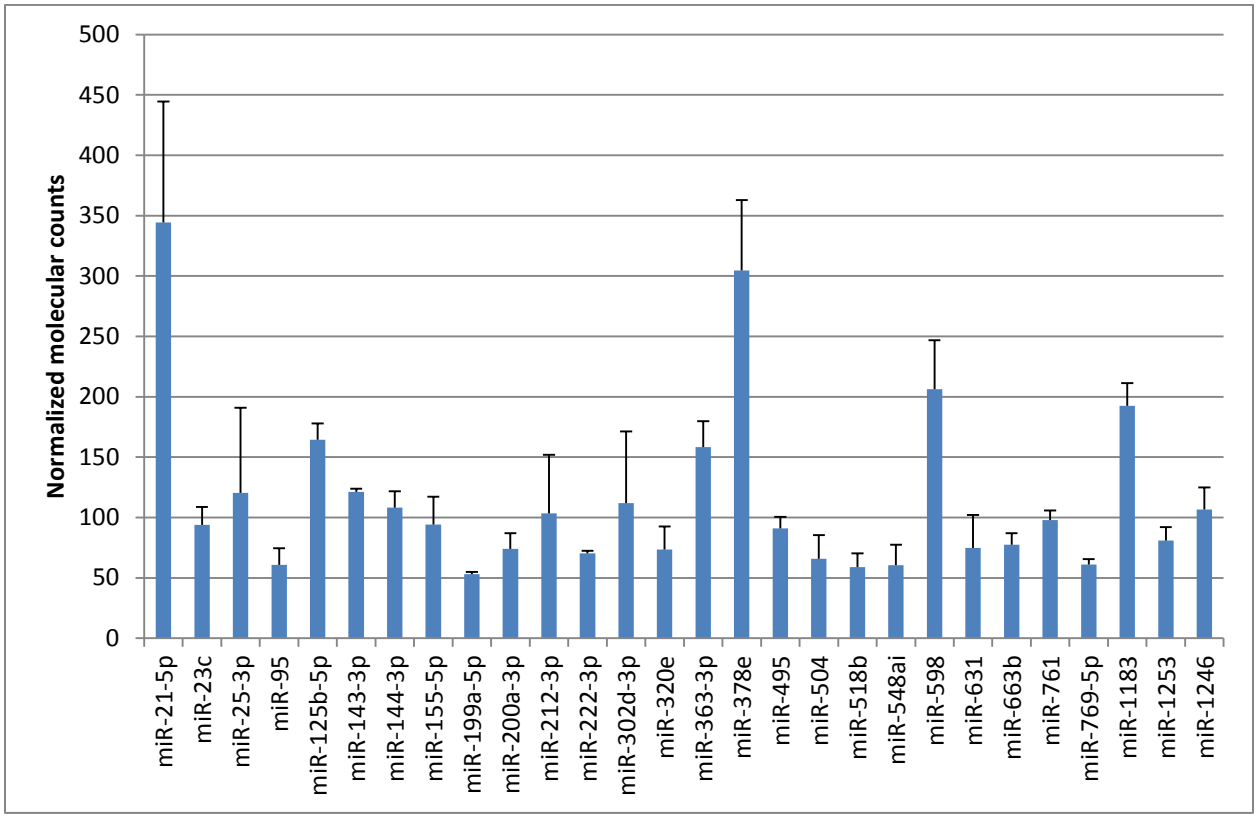

Normalized molecular counts
500
450
400
350
300
250
200
150
100
50
0
miR-21-5p
miR-23c
miR-25-3p
miR-95
miR-125b-5p
miR-143-3p
miR-144-3p
miR-155-5p
miR-199a-5p
miR-200a-3p
miR-212-3p
miR-222-3p
miR-302d-3p
miR-320e
miR-363-3p
miR-378e
miR-495
miR-504
miR-518b
miR-548ai
miR-598
miR-631
miR-663b
miR-761
miR-769-5p
miR-1183
miR-1253
miR-1246

Figure 8.

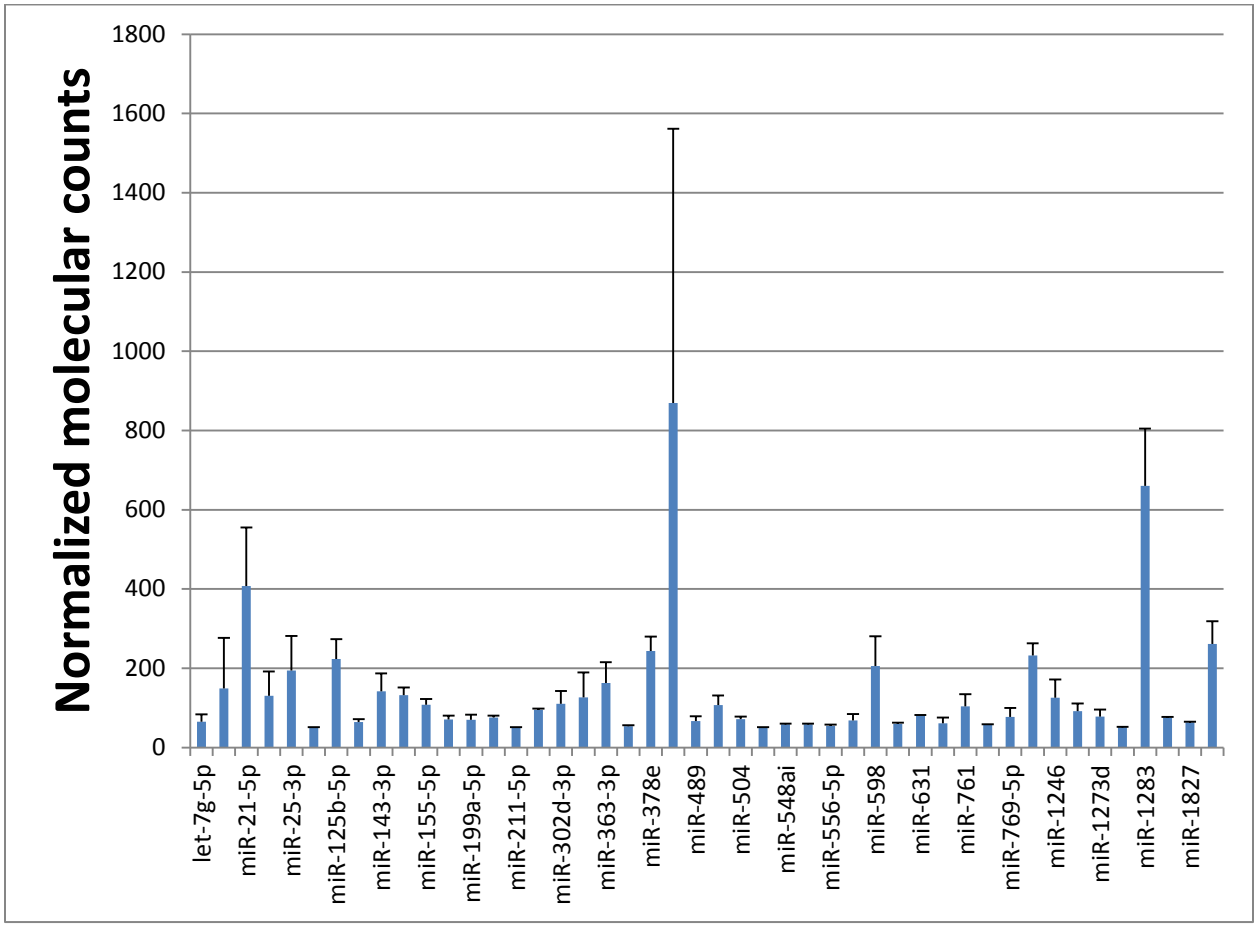

Normalized molecular counts
1800
1600
1400
1200
1000
800
600
400
200
0
let-7g-5p
miR-21-5p
miR-25-3p
miR-125b-5p
miR-143-3p
miR-155-5p
miR-199a-5p
miR-211-5p
miR-302d-3p
miR-363-3p
miR-378e
miR-489
miR-504
miR-548ai
miR-556-5p
miR-598
miR-631
miR-761
miR-769-5p
miR-1246
miR-1273d
miR-1283
miR-1827

Figure 9.

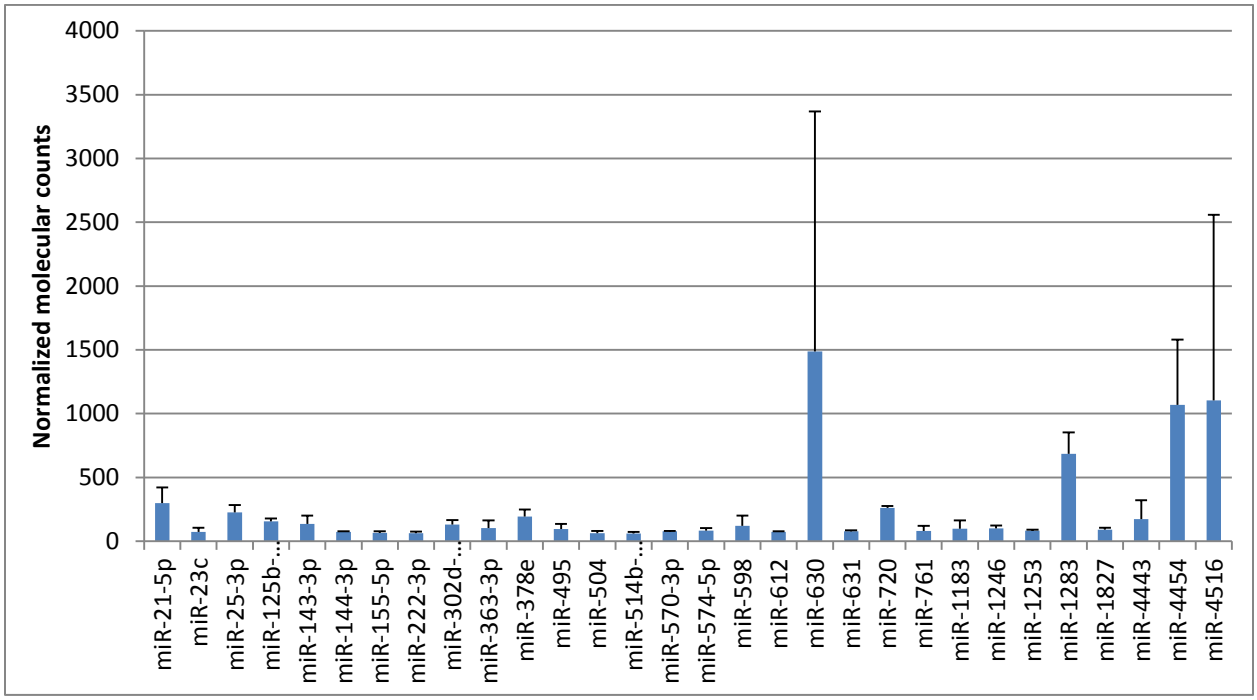

Normalized molecular counts
4000
3500
3000
2500
2000
1500
1000
500
0
miR-21-5p
miR-23c
miR-25-3p
miR-125b-...
miR-143-3p
miR-144-3p
miR-155-5p
miR-222-3p
miR-302d-...
miR-363-3p
miR-378e
miR-495
miR-504
miR-514b-...
miR-570-3p
miR-574-5p
miR-598
miR-612
miR-630
miR-631
miR-720
miR-761
miR-1183
miR-1246
miR-1253
miR-1283
miR-1827
miR-4443
miR-4454
miR-4516

Figure 10.

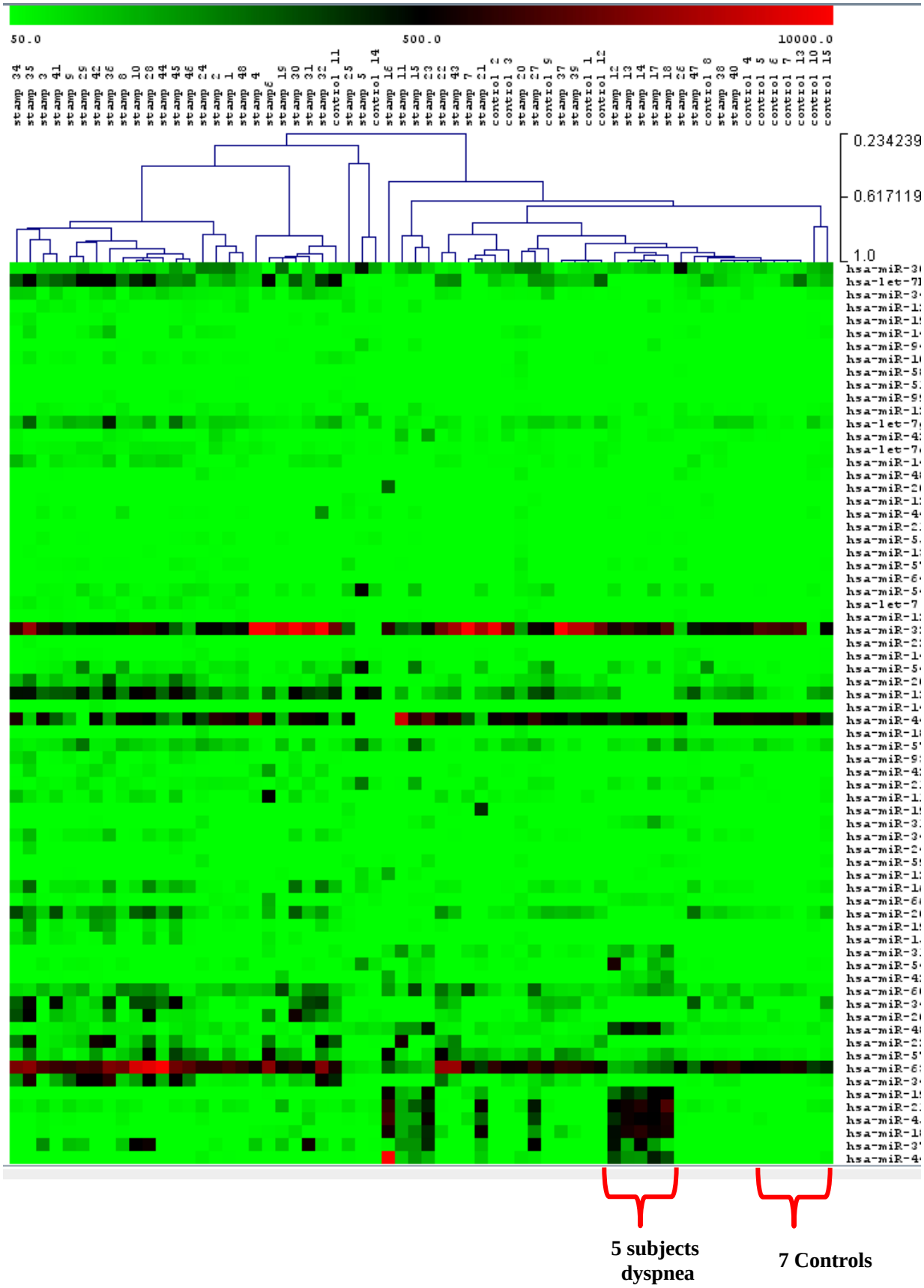

50.0
500.0
10000.0
0.234239
0.617119
1.0
miRNA 320e
miRNA 4516
miRNA 630
5 subjects dyspnea
7 Controls

Figure 11.

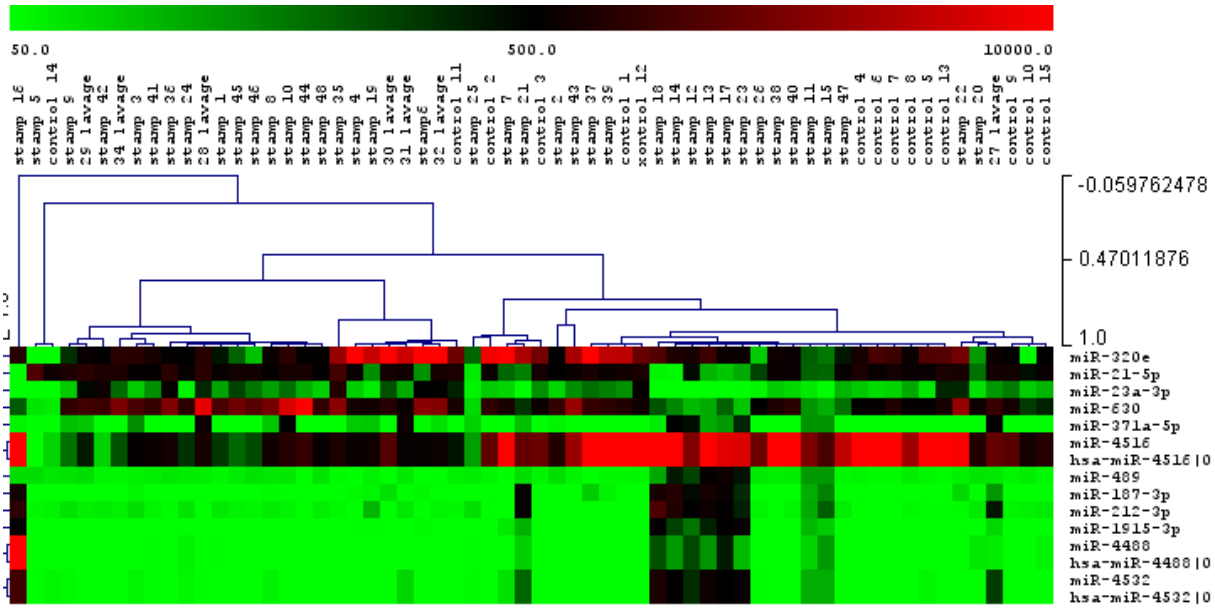

50.0
500.0
10000.0
-0.059762478
0.47011876
1.0
miR-320e
miR-21-5p
miR-23a-3p
miR-630
miR-371a-5p
miR-4516
hsa-miR-4516|0
miR-489
miR-187-3p
miR-212-3p
miR-1915-3p
miR-4488
hsa-miR-4488|0
miR-4532
hsa-miR-4532|0

Figure 12.

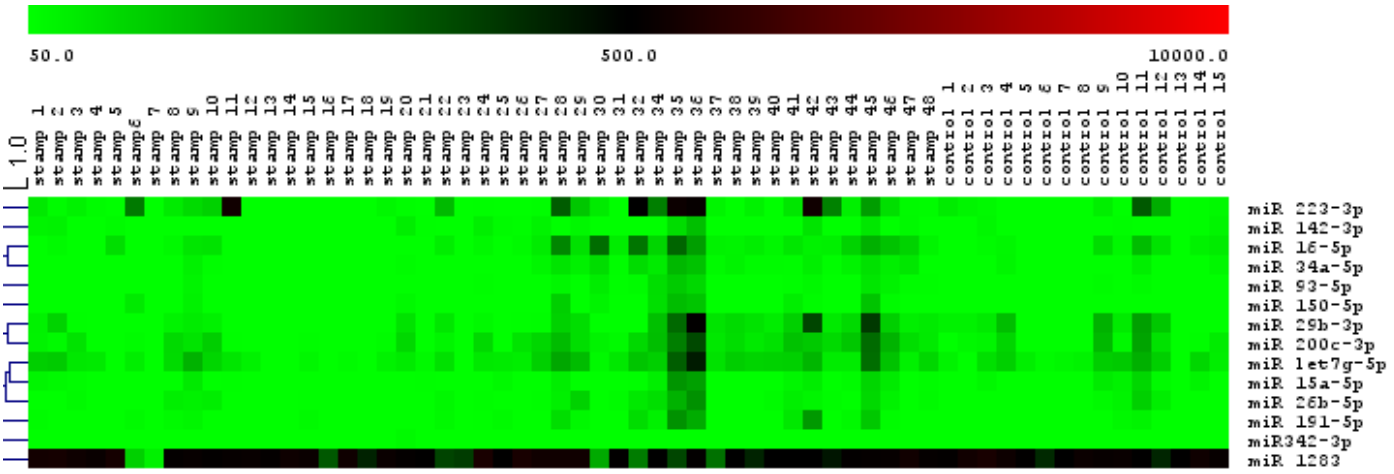

50.0
500.0
10000.0
1.0
stamp 1
stamp 2
stamp 3
stamp 4
stamp 5
stamp 6
stamp 7
stamp 8
stamp 9
stamp 10
stamp 11
stamp 12
stamp 13
stamp 14
stamp 15
stamp 16
stamp 17
stamp 18
stamp 19
stamp 20
stamp 21
stamp 22
stamp 23
stamp 24
stamp 25
stamp 26
stamp 27
stamp 28
stamp 29
stamp 30
stamp 31
stamp 32
stamp 34
stamp 35
stamp 36
stamp 37
stamp 38
stamp 39
stamp 40
stamp 41
stamp 42
stamp 43
stamp 44
stamp 45
stamp 46
stamp 47
stamp 48
control 1
control 2
control 3
control 4
control 5
control 6
control 7
control 8
control 9
control 10
control 11
control 12
control 13
control 14
control 15
miR 223-3p
miR 142-3p
miR 16-5p
miR 34a-5p
miR 93-5p
miR 150-5p
miR 29b-3p
miR 200c-3p
miR let7g-5p
miR 15a-5p
miR 26b-5p
miR 191-5p
miR342-3p
miR 1283

Figure 13.

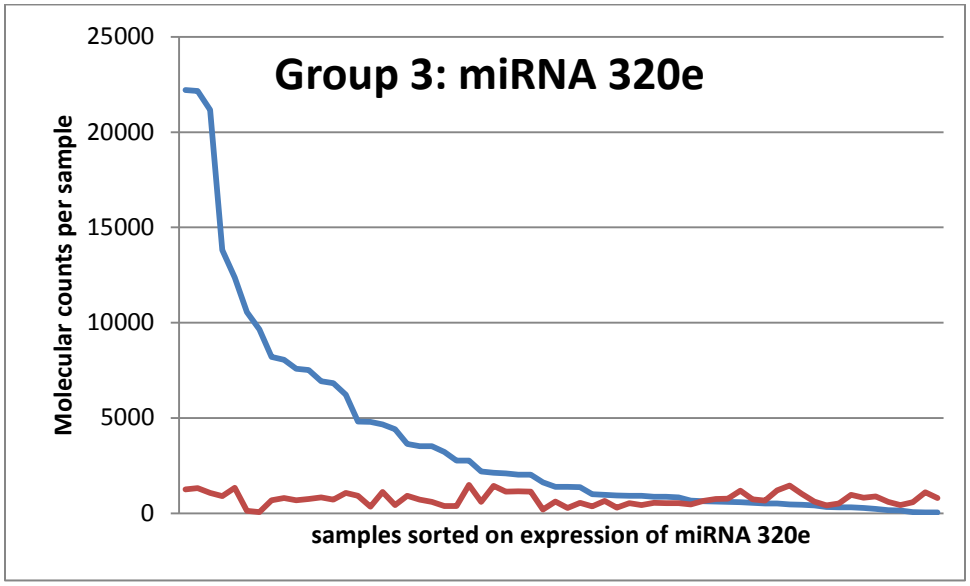


Figure 14.

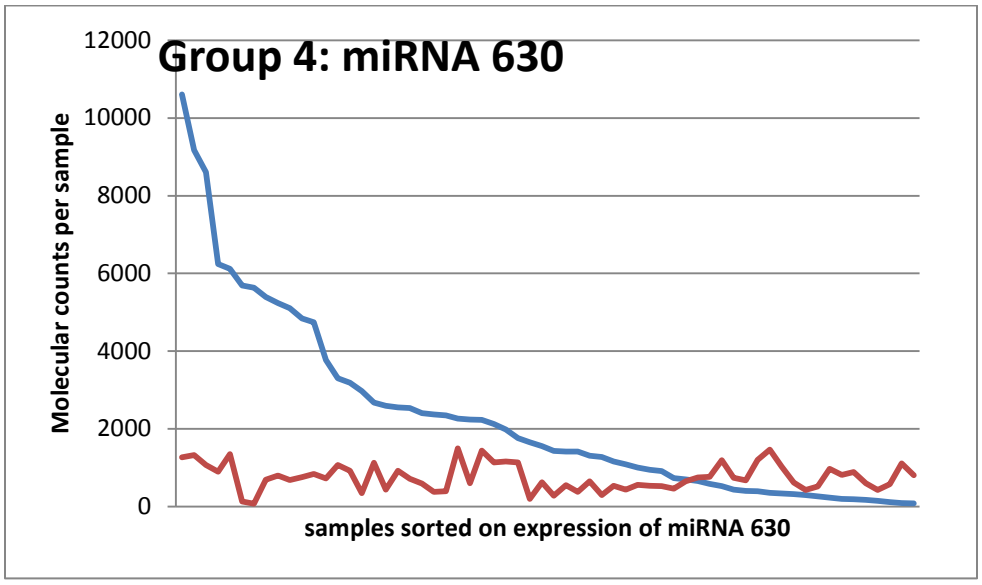


Figure. 15.

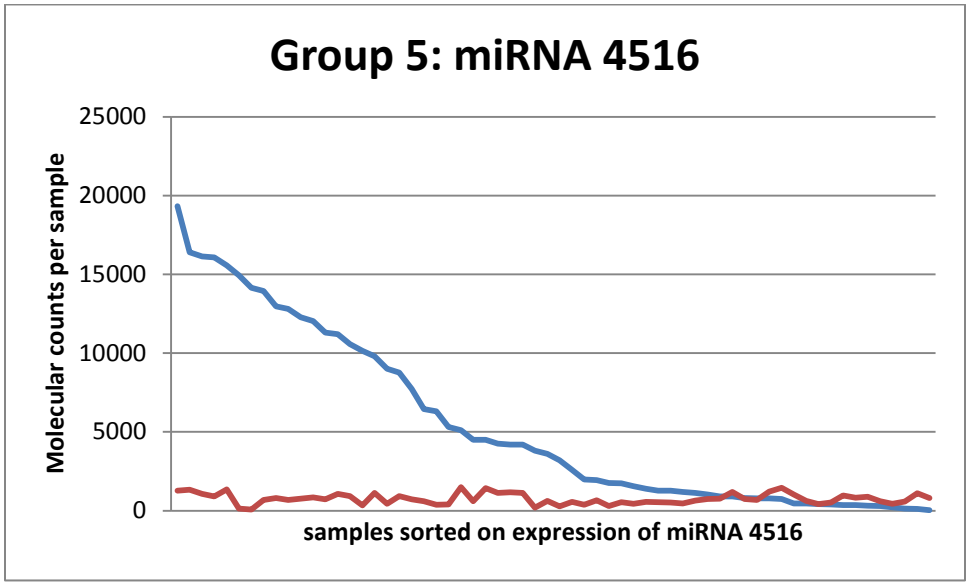

Group 5: miRNA 4516
25000
20000
15000
10000
5000
0
Molecular counts per sample
samples sorted on expression of miRNA 4516

Figure 16 A.

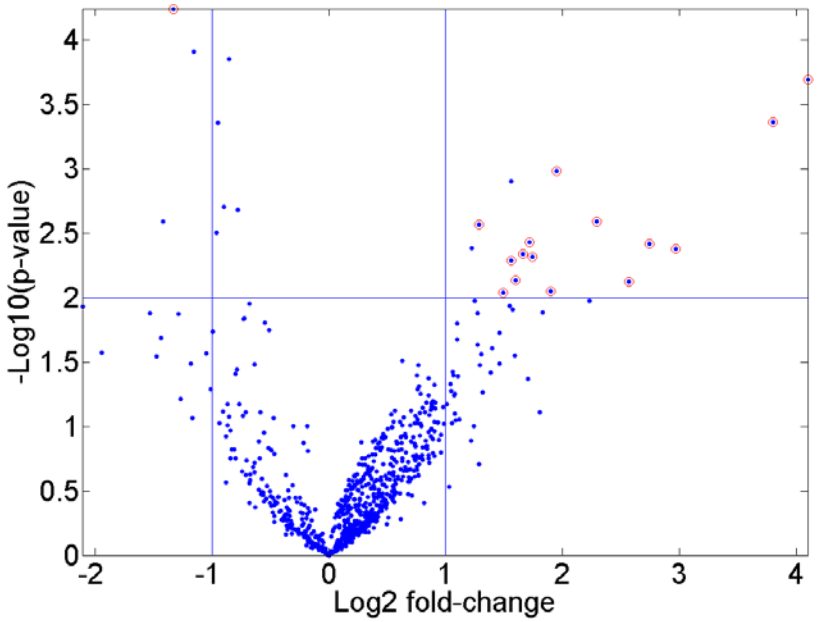


Figure 16 B.

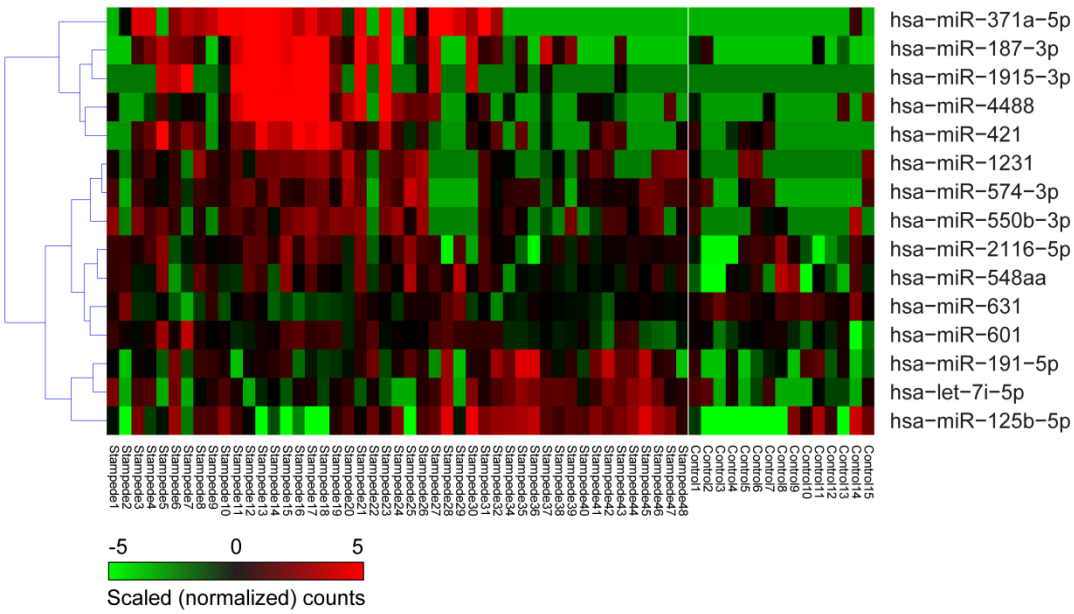
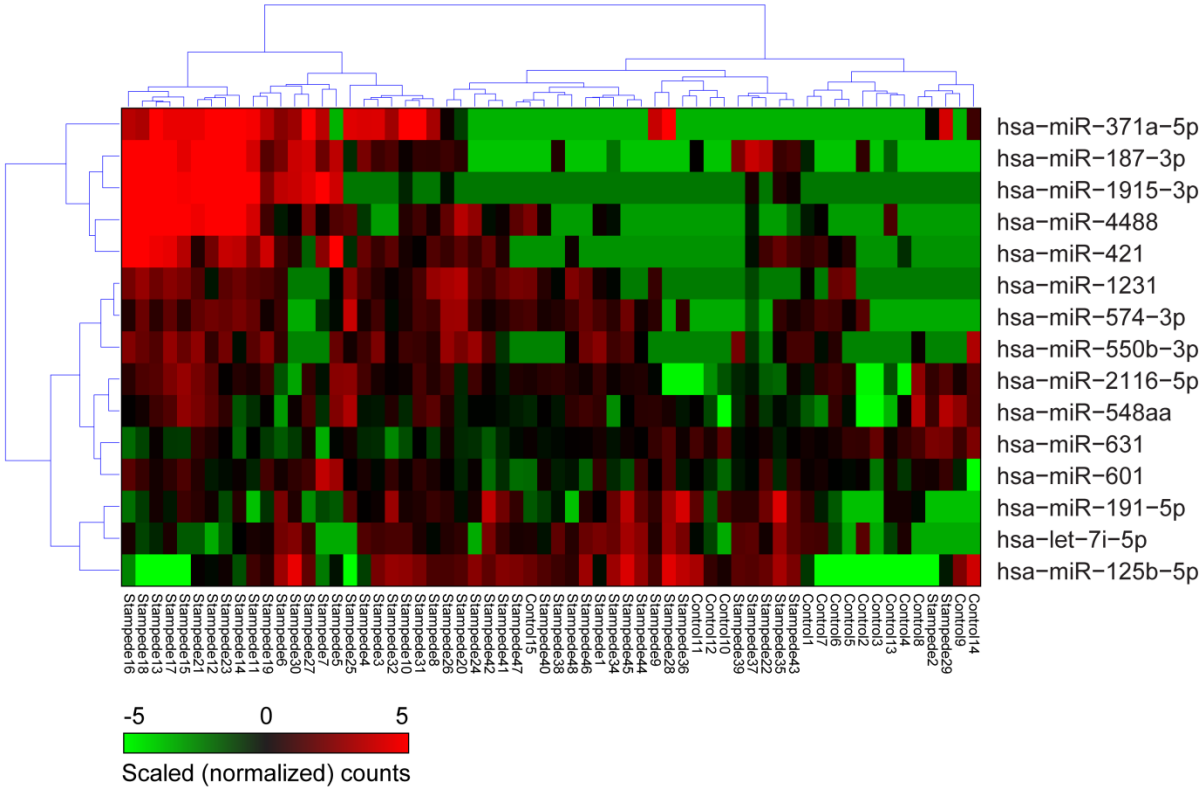
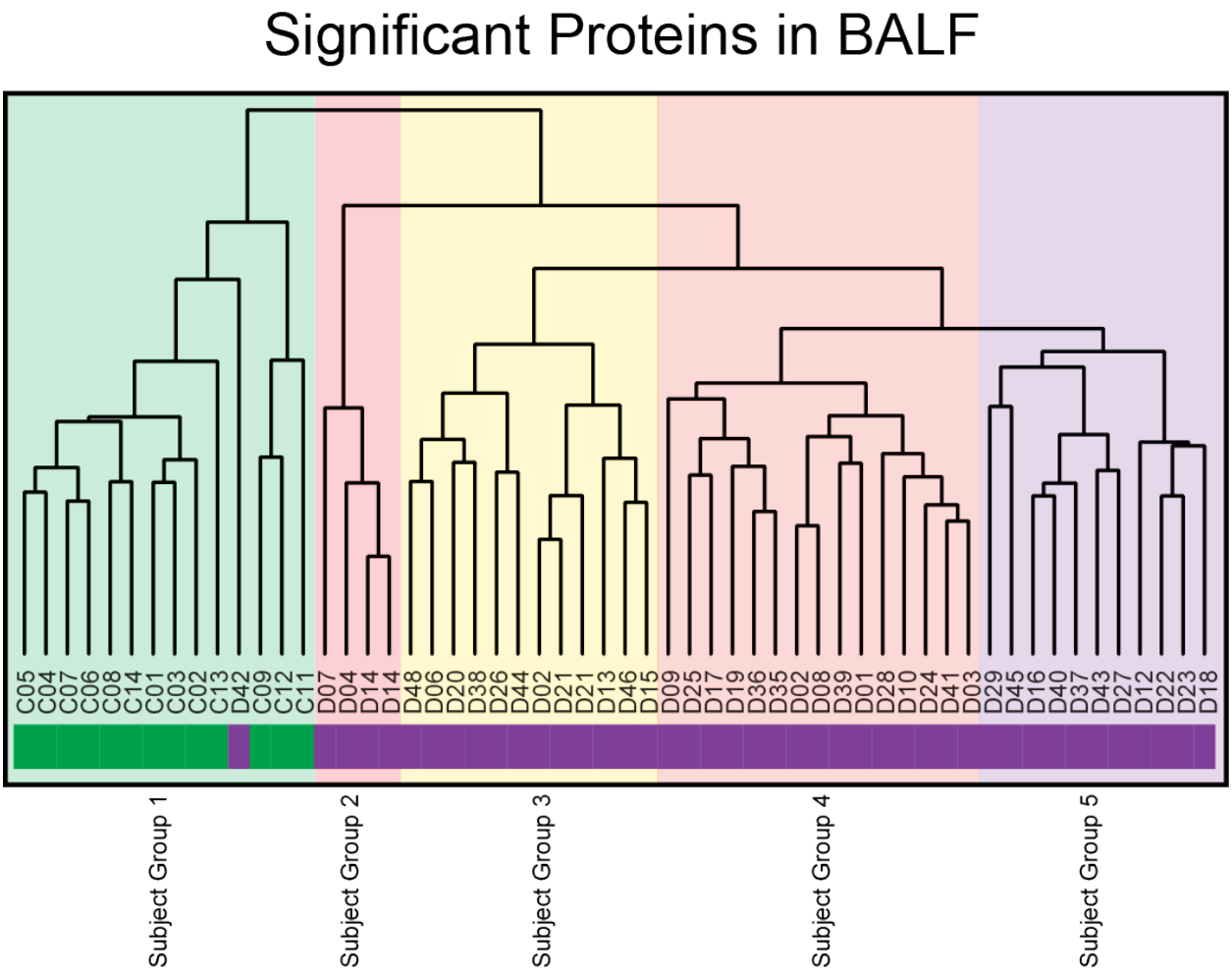


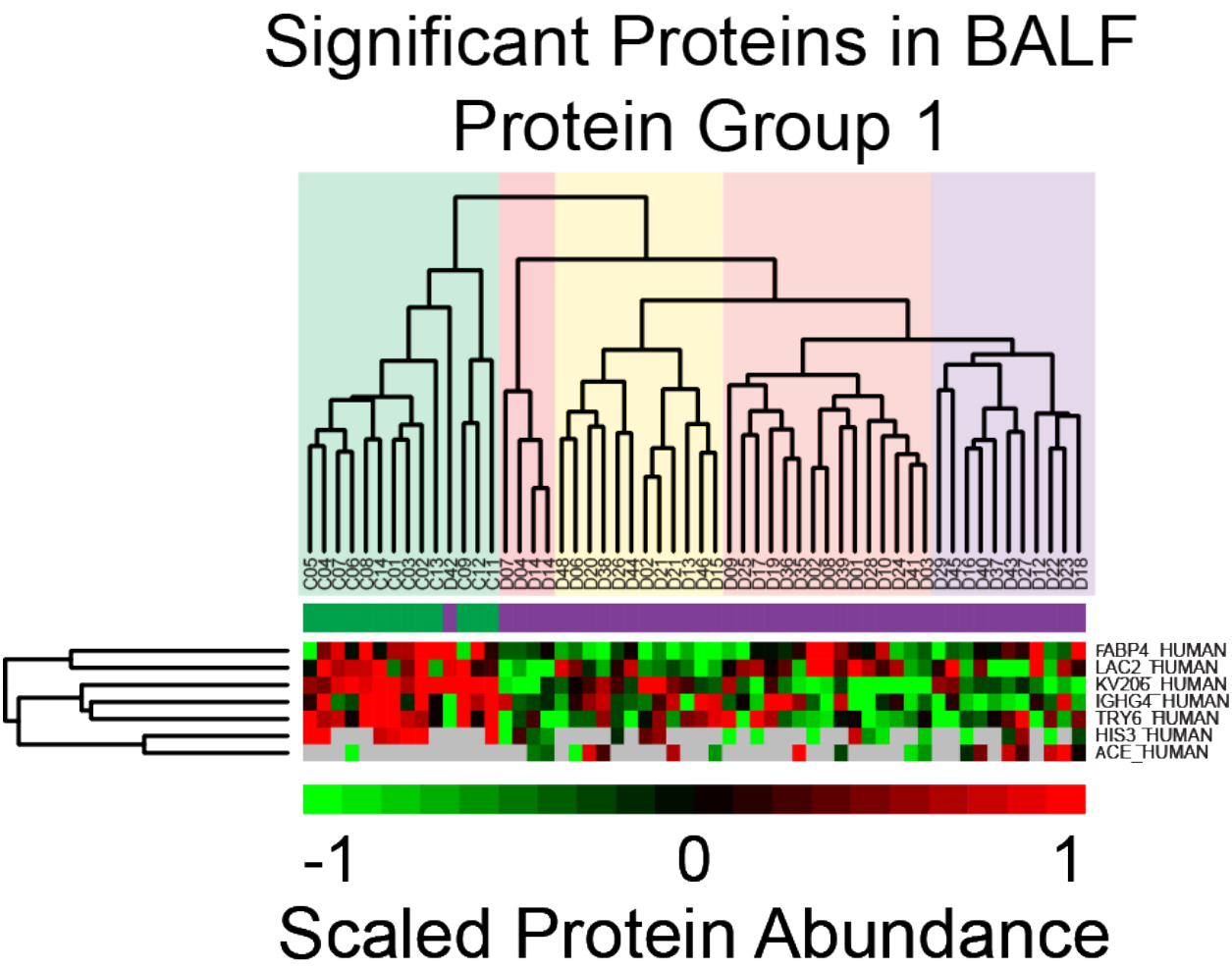
Figure 16 C.


hsa-miR-371a-5p
hsa-miR-187-3p
hsa-miR-1915-3p
hsa-miR-4488
hsa-miR-421
hsa-miR-1231
hsa-miR-574-3p
hsa-miR-550b-3p
hsa-miR-2116-5p
hsa-miR-548aa
hsa-miR-631
hsa-miR-601
hsa-miR-191-5p
hsa-let-7i-5p
hsa-miR-125b-5p
-5
0
5
Scaled (normalized) counts

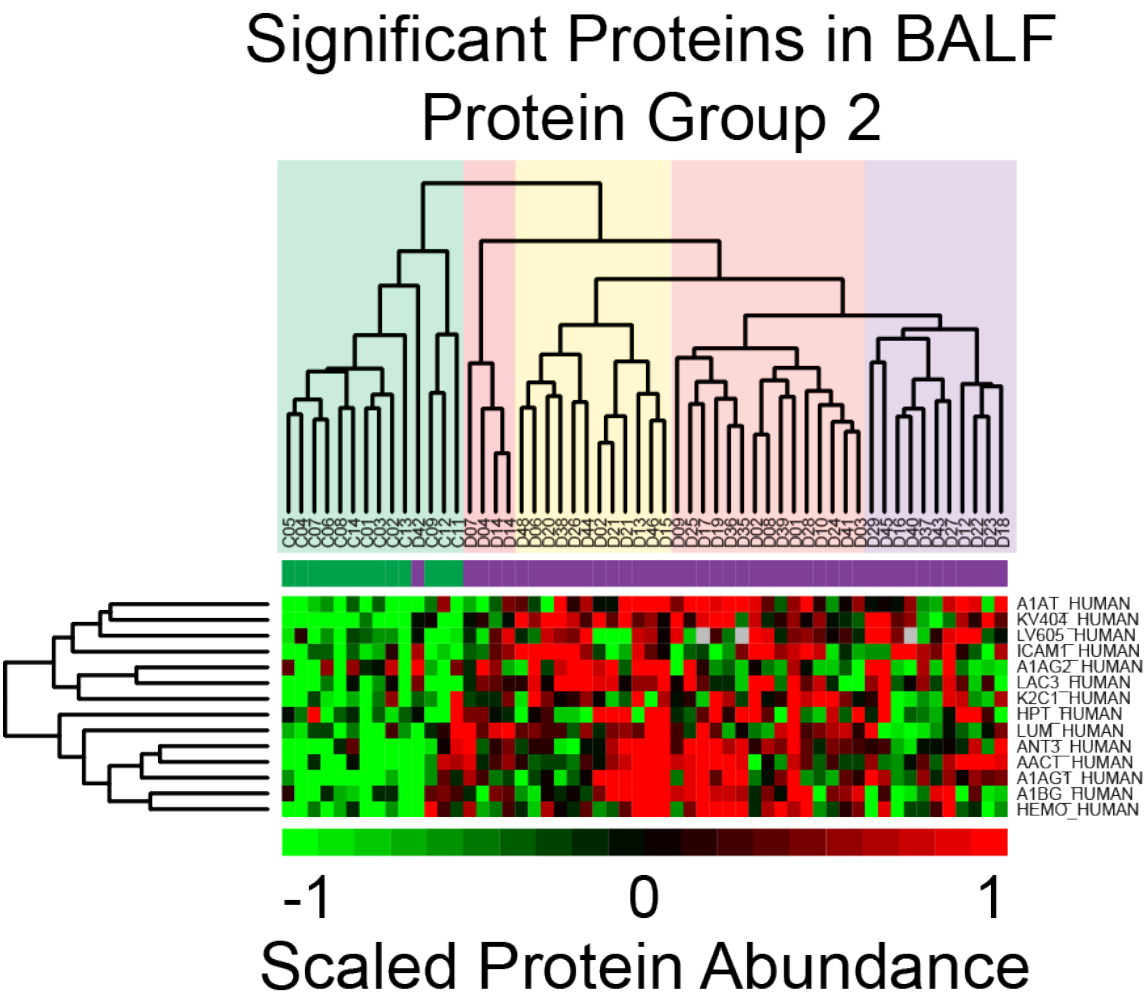
Supplementary Figure 1.



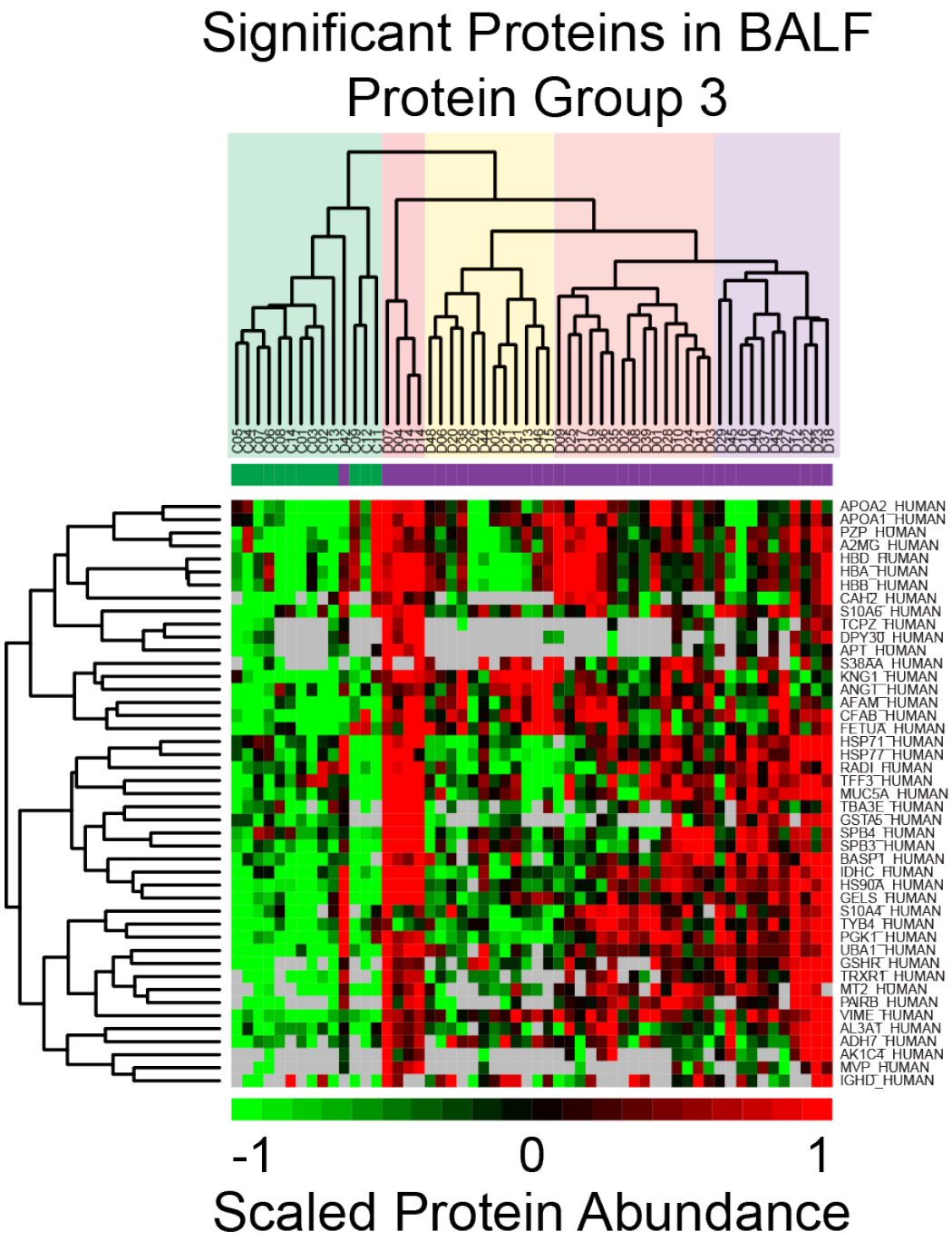
Supplementary Figure 2.



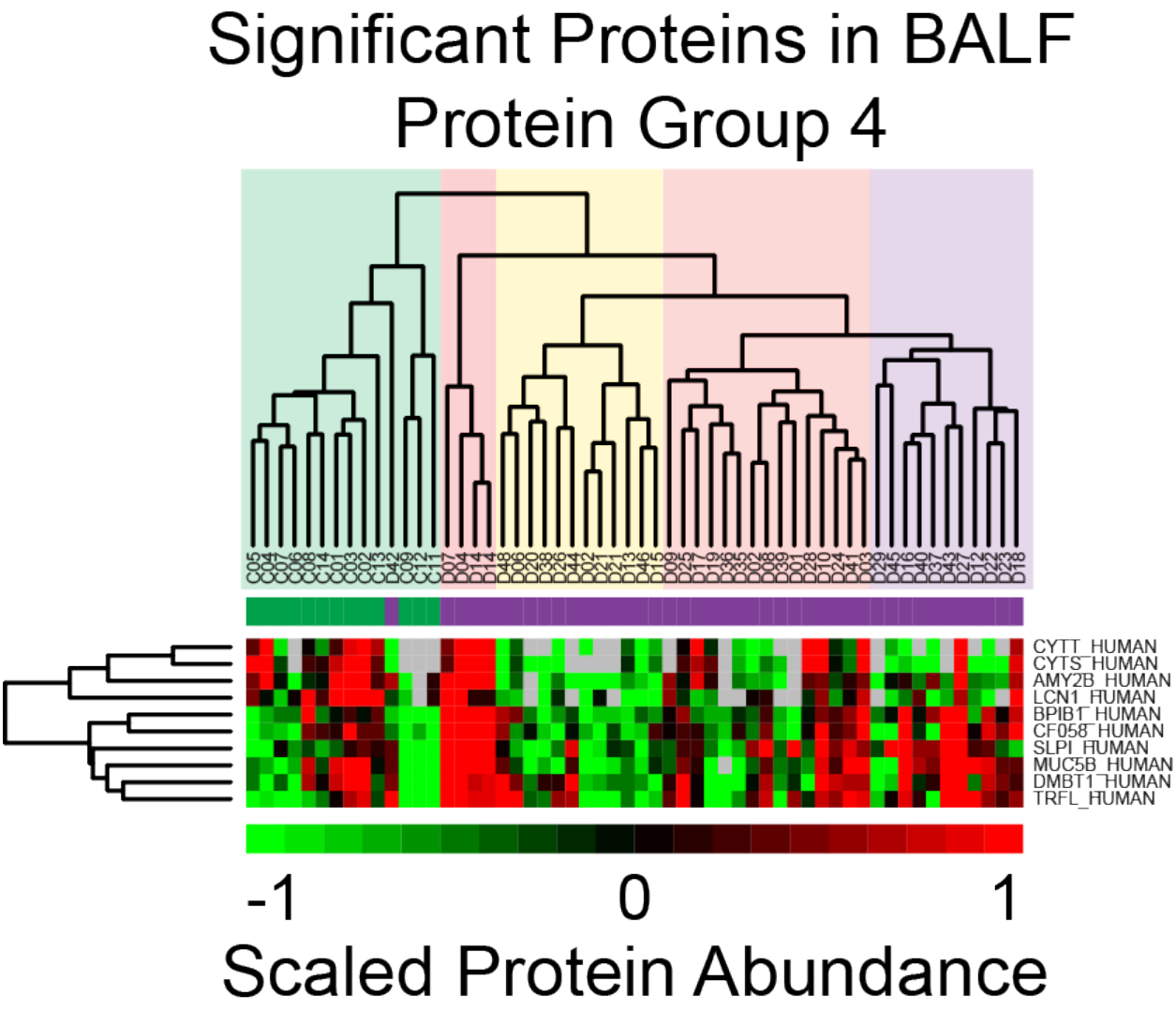
Supplementary Figure 3.



Supplementary Figure 4.



Supplementary Figure 5.



Supplementary Figure 6.